Leveraging chromatin accessibility for transcriptional regulatory network inference in T Helper 17 Cells

Emily R. Miraldi[1,2,10], Maria Pokrovskii[3], Aaron Watters[4], Dayanne M. Castro[5], Nicholas De Veaux[4], Jason A. Hall[3], June-Yong Lee[3], Maria Ciofani[6], Aviv Madar[7], Nick Carriero[4], Dan R. Littman[3,8,10], Richard Bonneau[4,5,9,10]

[1]Divisions of Immunobiology and Biomedical Informatics, Cincinnati Children's Hospital, Cincinnati, OH 45229, U.S.A.
[2]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45257, U.S.A.
[3]Molecular Pathogenesis Program, The Kimmel Center for Biology and Medicine of the Skirball Institute, New York, NY 10016, U.S.A.
[4]Center for Computational Biology, Flatiron Institute, New York, NY 10010, U.S.A.
[5]Department of Biology, New York University, New York, NY 10012, U.S.A.
[6]Department of Immunology, Duke University School of Medicine, Durham, NC 27710, U.S.A.
[7]Current Address: Novartis Institutes for Biomedical Research, Cambridge, MA 02139, U.S.A.
[8]The Howard Hughes Medical Institute
[9]Center for Data Science, New York University, New York, NY, 10010
[10]Corresponding Authors: emily.miraldi@cchmc.org, dan.littman@med.nyu.edu, rbonneau@flatironinstitute.org

Key words: transcriptional regulatory network, gene regulatory network, Th17, ATAC-seq, LASSO, StARS

Running Title: Genome-scale Transcriptional Models from ATAC-seq

Abstract:

Transcriptional regulatory networks (TRNs) provide insight into cellular behavior by describing interactions between transcription factors (TFs) and their gene targets. The Assay for Transposase Accessible Chromatin (ATAC)-seq, coupled with transcription-factor motif analysis, provides indirect evidence of chromatin binding for hundreds of TFs genome-wide. Here, we propose methods for TRN inference in a mammalian setting, using ATAC-seq data to influence gene expression modeling. We rigorously test our methods in the context of T Helper Cell Type 17 (Th17) differentiation, generating new ATAC-seq data to complement existing Th17 genomic resources (plentiful gene expression data, TF knock-outs and ChIP-seq experiments). In this resource-rich mammalian setting, our extensive benchmarking provides quantitative, genome-scale evaluation of TRN inference combining ATAC-seq and RNA-seq data. We refine and extend our previous Th17 TRN, using our new TRN inference methods to integrate all Th17 data (gene expression, ATAC-seq, TF KO, ChIP-seq). We highlight new roles for individual TFs and groups of TFs ("TF-TF modules") in Th17 gene regulation. Given the popularity of ATAC-seq, which provides high-resolution with low sample input requirements, we anticipate that application of our methods will improve TRN inference in new mammalian systems, especially *in vivo*, for cells directly from humans and animal models.

Introduction

Advances in genome-scale measurement and mathematical modeling herald opportunities for high-quality reconstruction of transcriptional regulatory networks (TRNs). TRNs describe the control of gene expression patterns by transcription factors (TFs) (Hecker et al. 2009; Chai et al. 2014; Marbach et al. 2012), providing mechanistic (and often genome-wide) insight into the complex regulation of cellular behavior (Bonneau et al. 2007). Measurements of chromatin state represent one such advance for TRN inference. For example, chromatin-immunoprecipitation with DNA microarray (ChIP-chip (Ren et al. 2000)) or sequencing (ChIP-seq (Robertson et al. 2007)) enable identification of an individual transcription factor (TF)'s binding sites genome-wide. ChIP experiments provide evidence for TF-target-gene interactions based on proximity of the TF binding site to the gene locus and have proved valuable for TRN inference (Lee et al. 2002; Ouyang et al. 2009; Ciofani et al. 2012). However, their utility is limited in cell types and settings (e.g., physiological) where sample material and/or a priori knowledge of key transcriptional regulators is scarce.

Genome-scale chromatin accessibility measurement (by Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE)-seq (Giresi et al. 2007), DNase1 hypersensitivity (DHS) (e.g., DNase1-seq) (Xi et al. 2007; Boyle et al. 2008), or most recently Assay for Transposase Accessible Chromatin (ATAC)-seq (Buenrostro et al. 2013)) highlight regulatory regions of the genome that are accessible for TF binding. Similarly, ChIP-seq for histone marks (Barski et al. 2007) (e.g., H3K27Ac, H3Kme3, etc.) correlate with promoters, enhancers, and/or other locus control regions, bound by TFs. Chromatin accessibility and histone-mark ChIP-seq measurements can partially overcome limitations in a priori knowledge of cell-type-specific TF regulators, if integrated with TF DNA-binding motifs (Pique-Regi et al. 2011). Large-scale efforts to characterize individual TF motifs are ongoing, with motifs currently available for ~700 TFs in human or mouse (~40% coverage for known TFs) (Weirauch et al. 2014; Jolma et al. 2010). Thus, chromatin state experiments integrated with TF analysis provide indirect DNA-binding evidence for hundreds of TFs, the scale of which would be impractical to attain with individual TF ChIP-seq experiments. Of techniques available, ATAC-seq best overcomes limitations in sample material availability, requiring two orders of magnitude fewer cells than a typical ChIP-seq, FAIRE-seq, or DHS experiment in standard, widely-adopted protocols. ATAC-seq is also possible at the resolution of a single cell (Buenrostro et al. 2015).

In the context of TRN inference, chromatin state measurements provide an initial set of putative TF-gene interactions, based on evidence of TF binding near a gene locus; evidence can be direct (TF ChIP-seq) or indirect (e.g., TF motif occurrence in accessible chromatin) and can be used to refine gene expression modeling (Blatti et al. 2015; Wilkins et al. 2016; Qin et al. 2014). Integration of chromatin state data in TRN inference could mitigate false-positive and false-negative TF-gene interactions expected from chromatin-state data analyzed in isolation (Äijö et Bonneau 2016; Siahpirani et Roy 2016). Sources of false positives and negatives include the following: (1) non-functional binding, (2) long-range interactions between genes and regulatory regions (3) the limited availability of individual TF ChIP experiments and incomplete knowledge of TF DNA-binding motifs, and (4) non-bound accessible motifs. Thus, an initial TRN derived solely from chromatin state data can be considered a useful but noisy prior, to be integrated with other data types for TRN inference.

Genome-scale inference of TRNs in mammalian settings is an outstanding challenge, given the increased complexity of transcriptional regulatory mechanisms relative to simpler eukaryotes. Thus, chromatin state measurements are especially important for mammalian TRN inference. Construction of a genome-scale TRN for *in vitro* T Helper Cell Type 17 (Th17) differentiation

provided a proof-of-concept for this idea (Ciofani et al. 2012). Rich genomics datasets informed the Th17 TRN: 156 RNA-seq experiments (including knock out (KO) of 20 TFs), ChIP-seq of 9 key Th17 TFs, and microarray from the Immunological Genome Project (Heng et al. 2008). We used the *Inferelator* algorithm (Bonneau et al. 2006; Madar et al. 2010) to infer TRNs from the RNA-seq and microarray data, and independent methods to build networks from TF ChIP and KO data. We showed that rank combination of the networks performed best recovering known Th17 genes and GWAS disease genes associated with Th17 pathologies.

Since our previous publication of the Th17 TRN, the *Inferelator* algorithm underwent several key developments that improve inference in unicellular organisms and are expected to improve TRN inference in a mammalian setting (Greenfield et al. 2013a; Arrieta-Ortiz et al. 2015). While the Inferelator's core model of transcriptional regulation still describes differential gene expression as a sparse multivariate linear function of TF activities, the methods used to solve for the TF-gene interaction terms as well as techniques for TF activity estimation have changed. For example, the current version of the Inferelator uses a Bayesian approach to incorporate prior information (G-priors with Bayesian Best Subset Regression (BBSR)) (Greenfield et al. 2013b). This framework has several desirable features. For example, a TF-gene edge in the prior will not be incorporated in the final TRN without support from the gene expression data, and new edges (not included in the prior) can be learned if there is sufficient support. The current "*Inferelator*-BBSR" presents a promising means to leverage a noisy, chromatin-state-derived prior network in a mammalian setting.

The focus of this work is development of mammalian TRN inference methods from chromatin accessibility and gene expression, data types available or likely feasible for an ever-growing number of cell types and biological conditions. In the context of mammalian TRN inference, several studies build TRNs directly from chromatin accessibility without further refinement by multivariate gene expression modeling (Neph et al. 2012; Rendeiro et al. 2016). Several other studies leverage variance in paired RNA-seq and ATAC-seq datasets; these TRN methods are exciting developments but require that ATAC-seq data be available for all or most RNA-seq conditions (Karwacz et al. 2017; Ramirez et al. 2017; Duren et al. 2017). In contrast, the present work is geared for TRN inference from RNA-seq and ATAC-seq, where ATAC-seq need not exist for more than one gene expression condition.

Development of any TRN inference method requires a comprehensive benchmark with a realistic experimental design, a recurrent challenge in computational biology. To address this challenge and enable rigorous benchmarking of our methods, we extended the genomics resources previously developed by our labs for Th17 TRN inference, using the same Th17 differentiation protocol to generate new ATAC-seq samples. In addition, we augment our initial set of 153 RNA-seq experiments with additional (including unpublished) RNA-seq experiments for a total of 254 *in vitro* Th17 and other T Helper (Th) Cell RNA-seq experiments. Using the RNA-seq and ATAC-seq as input to TRN methods, the quality of resulting TRNs is quantified with precision-recall of "gold-standard" TF-gene interactions in Th17 cells (supported by TF KO and ChIP data (Ciofani et al. 2012; Yosef et al. 2013)). As an alternative to performance relative to a gold standard (unavailable for most cell types), we test TRN quality with out-of-sample gene expression prediction. After extensive benchmarking and method development, we take what we have learned and infer an updated TRN model for Th17 cells from all available data (RNA-seq, ATAC-seq, ChIP-seq and TF KO network). Thus, in addition to methods and a genomics resource for mammalian TRN construction, we provide new predictions about transcriptional regulation in Th17 cells, easily accessible to the community in an interactive Jupyter notebook framework.

Results

Construction of Th17 benchmark for TRN inference from ATAC-seq and RNA-seq

To test the feasibility of TRN inference from chromatin accessibility and gene expression alone, we generated an ATAC-seq dataset in Th17 cells and other *in vitro* polarized T cells, matching a subset of experimental conditions and time points from the original publication (Ciofani et al. 2012) (**Figure 1A**). We used Principal Component Analysis (PCA) to determine sources of variance in the chromatin accessibility profiles (**Figure 1A**) (see **Methods**). Time was the most important driver of chromatin accessibility patterns; the first principal component (PC) explained 55% of the variance and captured accessible regions (peaks) that changed from 0hr (naïve CD4+ T cells) to 48hrs (Th17, Th0, Th2, Treg). The second PC captured accessibility differences between Th17 and the other T cell polarization programs. The Th17 ATAC-seq dataset also includes several perturbation conditions, including TF KO of *Stat3* and *Maf* for Th17 and Th0 conditions (48hr). STAT3 is required for Th17 differentiation and *Stat3* KO dramatically altered Th17 chromatin accessibility, leading to a Th0-like profile (**Figure 1A**, red arrow), while *Maf* KO clustered with Th17 (**Figure 1A**, gray arrow). Addition of other Th17-polarizing cytokines (IL-1β, IL-23, and/or IL-21) to the Th17 differentiation media (TGF-β and IL-6) did not dramatically change chromatin accessibility patterns at 48 hours.

To the 153 RNA-seq experiments from the original publication, we added an additional 101 RNA-seq experiments, both published (13) and unpublished (88), for a total of 254 experiments (**Figure 1B**, **Methods**). The majority of the samples (166) were Th17, spanning 1hr to 96hrs, involving KO, siRNA knock down, and/or drug inhibitors of TFs and signaling molecules. The study design also included other T cell polarizations (Th0 (53), Tr1 (9), and Th1/Th2/Treg (2 each)) as well as naive CD4 T cells (25). Mirroring chromatin accessibility patterns, PCA of the gene expression data revealed time and T Cell polarization conditions to be important drivers of transcriptional variation (**Figure 1B**, **Figure S1**).

While gene expression data are the only required input for the *Inferelator* algorithm (**Methods**), we hypothesized that the inclusion of ATAC-seq data could improve TRN inference. As described above, integration of ATAC-seq with TF motifs can provide indirect evidence for TF-binding events associated with altered chromatin state (**Figure 1A**) and transcription (**Figure 1B**). In this study, we generated "prior" networks of TF-gene interactions by associating TFs with putative target genes based on TF motif occurrences within accessible regions near genes (e.g., +/-10 kb gene body, see **Methods**). **Table S1A** provides high-level statistics for two ATAC-seq-derived prior networks: the A(Th17) prior, limited to motif analysis of peaks accessible in 48h Th17 conditions only, and the A(Th) prior, which includes accessible regions from all samples in **Figure 1A**. The A(Th17) and A(Th) priors contained ~1.1 and ~2 million putative TF-gene interactions for ~800 TFs. These noisy priors help guide network structure for TRN inference (**Figure 1C**). Although we do not have an estimate of false edges in our ATAC-seq priors, the study design (**Figure 1C**) enables quantitative performance evaluation of the resulting TRNs. Specifically, the TRNs are evaluated based on precision and recall of TF-gene interactions in an independent gold standard (G.S.), composed of edges supported by TF KO and/or TF ChIP data. **Table S1A** provides statistics on the G.S. networks and their overlap with other prior information sources. As the precision-recall performance metric is limited to the TFs previously selected for KO (25 TFs) and/or ChIP (9 TFs) analysis, we also quantify TRN model performance based on out-of-sample gene expression prediction.

Using precision-recall and out-of-sample prediction metrics, we evaluate the effects of several key modeling decisions. **Figure 1C** outlines inputs to the *Inferelator* algorithm. From the gene expression dataset, we seek to model the expression patterns of 3578 "target" genes (**Methods**)

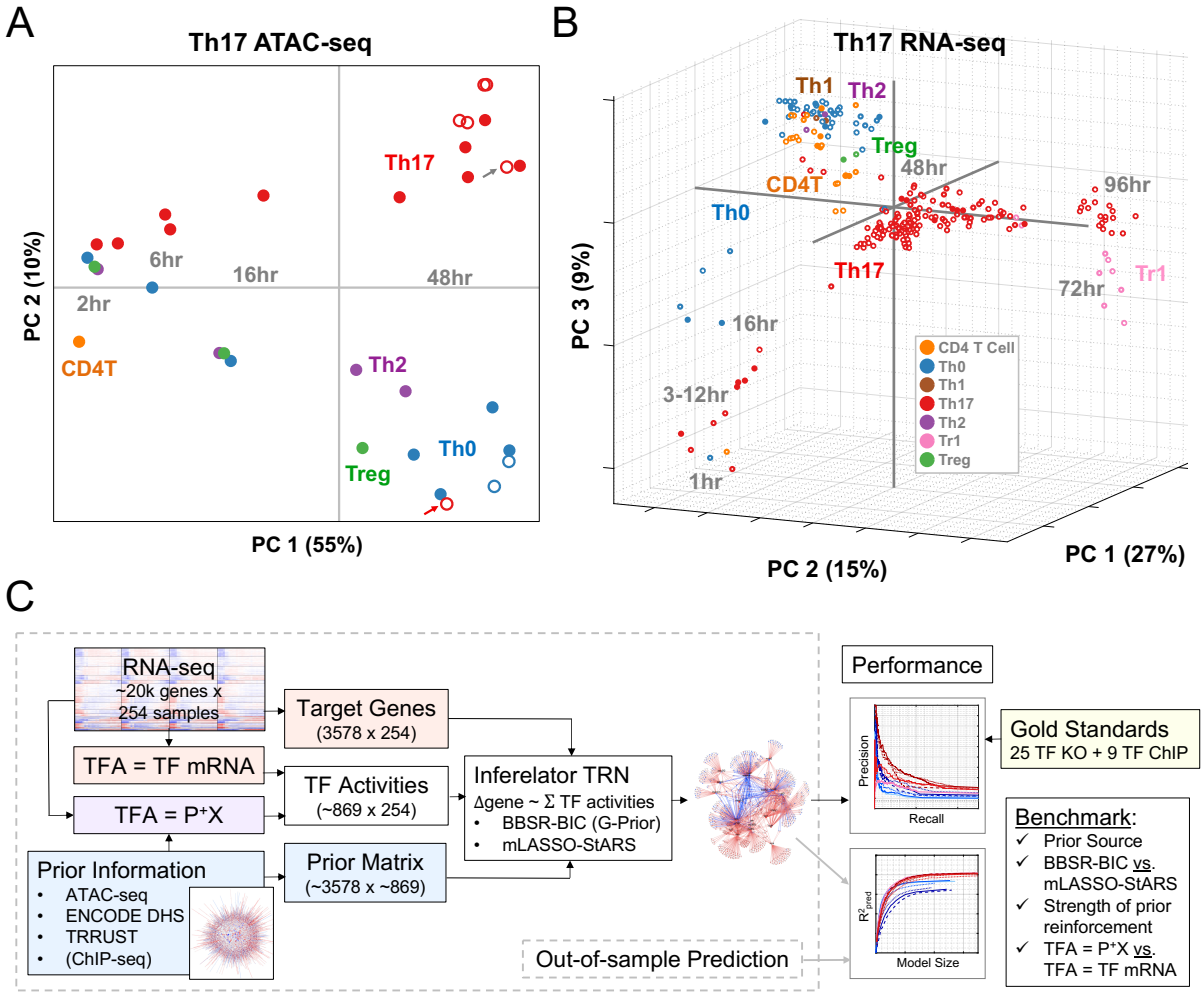

Figure 1. (A) PCA of Chromatin Accessibility Profiles. The 33 ATAC-seq samples are plotted as a function of robustly normalized ATAC-seq peak intensities in PCA space, using the reference set of 63,049 ATAC-seq peaks identified. Open circles denote experimental conditions that deviate from the standard T Cell differentiation conditions (e.g., genetic deletion, additional cytokines). Gray and red arrows indicate Maf and Stat3 KO Th17 conditions, respectively. **(B) PCA of Gene Expression Profiles.** The 254 RNA-seq samples are plotted as a function of all genes in PCA space. **(C) Study Design.** The gene expression data is used to construct both a target gene matrix and to estimate TFA, either with the prior (TFA=P$^+$X) or independently of the prior (TFA=TF mRNA). The *Inferelator* takes target gene and TFA matrices as input and builds models gene expression as a function of TFAs, using either BBSR-BIC or mLASSO-StARS methods. The prior matrix is an optional *Inferelator* input that can be used to reinforce TF-gene interactions with support from the prior. The quality of resulting TRNs is measured using (1) precision-recall of edges in the gold standard and (2) out-of-sample prediction error. The design enables quantitative comparison of prior information source, TFA estimation technique, *Inferelator* method, and strength of prior reinforcement.

as functions of TF activities (TFAs). Protein TFAs are rarely measured and technically infeasible for most TRN experimental designs. Thus, the TF activity can be considered a hidden (or latent) variable in the context of TRN inference (Liao et al. 2003; Fu et al. 2011). The most common estimate of TFA is TF mRNA. However, many TF transcriptional activities require protein post-translational modification, and, in those cases, TF mRNA can be a poor proxy for protein transcriptional activity. Thus, TFA estimation based on prior knowledge of TF target genes provides an alluring alternative, as it appears to be technically feasible – requiring only partial a priori knowledge of TF-gene interactions and gene expression data (see **Methods**). TFA

estimation based on target gene expression quantitatively improved TRN inference using highly curated, literature-based priors in unicellular organisms (Arrieta-Ortiz et al. 2015; Tchourine et al. 2017). However, "prior-based" TFA estimation (from ATAC-seq and other data sources) has yet to be quantitatively evaluated in a mammalian setting at scale.

We test two methods for model selection (i.e., TF-gene interaction terms to include in the TRN models): (1) the most recently published version of *Inferelator* (Bayesian Best Subset Regression with Bayesian Information Criteria for model selection (BBSR-BIC) (Arrieta-Ortiz et al. 2015)) and (2) an alternative proposed here, modified Least Absolute Shrinkage and Selection Operator (Studham et al. 2014; Gustafsson et al. 2015) with Stability Approach to Regularization Selection (Liu et al. 2010) (mLASSO-StARS), to overcome challenges inherent to mammalian TRN inference. Specifically, we hypothesized that mLASSO-StARS would (1) scale better with the increased transcriptional complexity expected in a mammalian setting (e.g., larger models) and (2) stability-based StARS model selection would perform better for relatively smaller mammalian gene expression datasets (10s-100s of samples versus 1000s of samples in the unicellular organisms for which Inferelator BBSR-BIC was originally developed) (detailed in **Methods**). Thus, we compare mLASSO-StARS and state-of-the-art BBSR-BIC for mammalian TRN inference (Arrieta-Ortiz et al. 2015). Notably, StARS model selection has been used in several related biological contexts already (Kurtz et al. 2015; Zhang et al. 2016; Caballe Mestres et al. 2017).

Prior information can enter our inference procedure at two steps (1) prior-based TFA estimation (described above) and (2) to reinforce prior-supported TF-gene interactions (edges) at the multivariate regression step, using either BBSR-BIC or mLASSO-StARS (**Figure 1C**, **Methods**). The strength of prior reinforcement controls the relative contribution of prior evidence (e.g., TF ChIP, ATAC-seq motif analysis) to evidence from the gene expression model (variance explained by individual TFs). Effectively, the strength of prior reinforcement determines the frequency with which prior edges enter the network relative to edges supported by gene expression modeling alone. The appropriate level of prior reinforcement should be a function of (1) the quality of the prior information and (2) quality and number of gene expression samples available for model building. Clearly, the strength of prior reinforcement will influence TRN inference; thus, we test several levels of reinforcement in our study design. We also test several sources of prior information, in addition to ATAC-seq.

Modified LASSO-StARS improves inference of a mammalian TRN

As illustrated in **Figure 1C**, we use precision and recall over multiple benchmark networks to quantitatively evaluate the impact of key modeling decisions on resulting Th17 TRN models (prior sources, method of TFA estimation, model selection method (BBSR-BIC or mLASSO-StARS), and strength of prior reinforcement). Both BBSR-BIC and mLASSO-StARS provide confidence estimates for every TF-gene interaction ("edge") predicted (**Methods**), and these confidences can be used to prioritize TF-gene interactions and rank edges. For each confidence cutoff, one can evaluate the set of predictions in terms of precision (fraction of predicted edges in the gold standard) and recall (fraction of the total gold standard recovered). There is a well-known trade-off between precision and recall. In general, a high confidence cutoff (yielding a small network) favors high precision but low recall, while a low confidence cutoff (many interactions included in the network) favors high recall at the cost of lower precision. Therefore, it is instructive to plot precision and recall as a function of confidence cutoffs, as in **Figure 2A**. Calculating the area under the precision-recall curve (AUPR) provides a useful summary of precision-recall curve.

The quality of precision-recall analysis is dependent on the quality of the gold standard (G.S.). In this study, we have access to 25 TF KO RNA-seq experiments (Th17 48h time point), from which TF-target gene interactions were derived from differential expression analysis (Ciofani et al. 2012;

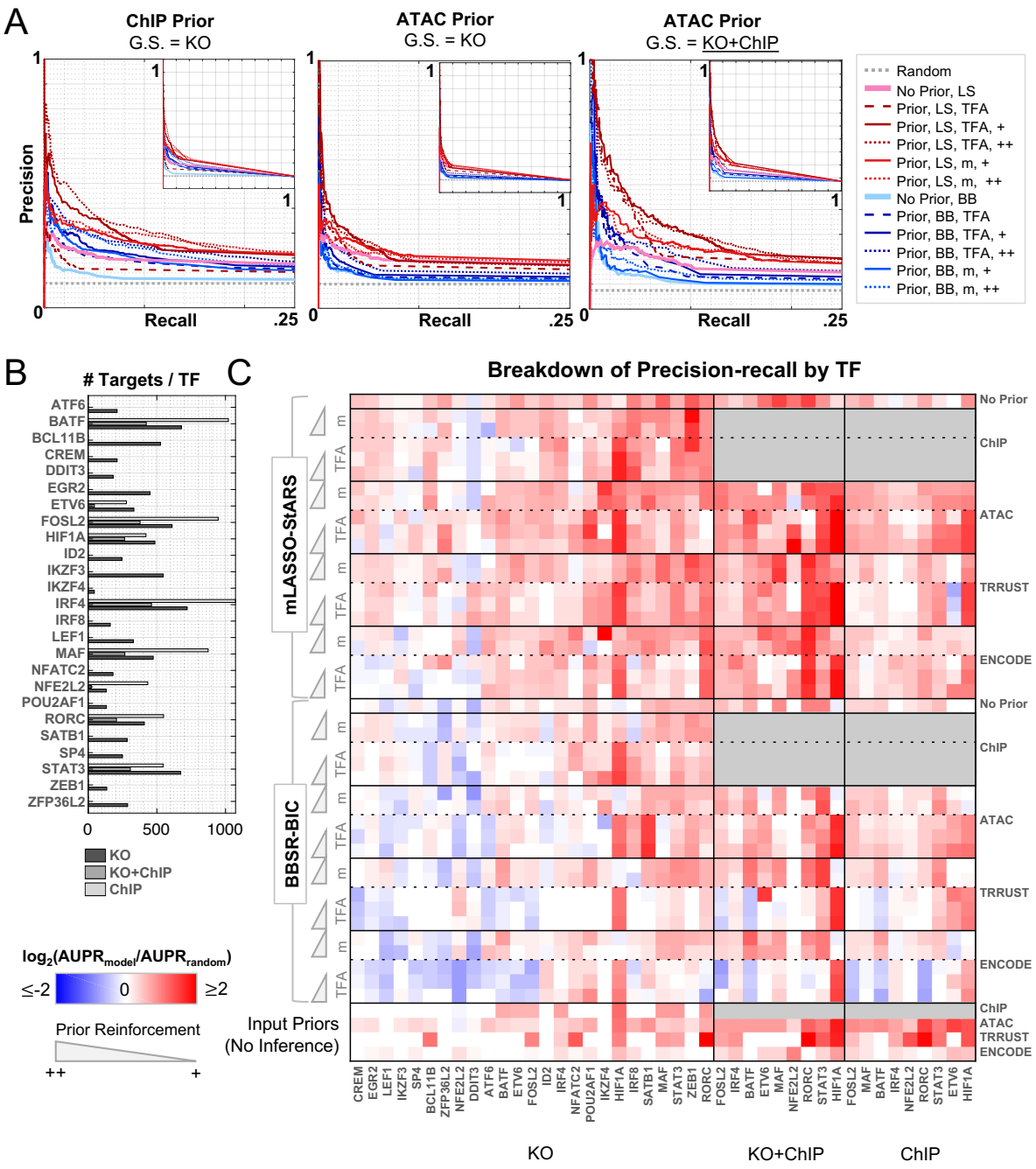

Figure 2. (A) Precision-recall of Th17 TRNs. The left two panels enable comparison of TRNs built from ChIP versus ATAC priors, quantified by precision-recall of the KO gold standard (G.S.) (25 TFs, 8875 interactions), insets display the full curve. The performance of several TRNs are plotted for each prior, based on Inferelator method (LS = mLASSO-StARS (reds), BB = BBSR-BIC (blues)), TFA estimate (m = TF mRNA, TFA = $P^{+}X$), and "+" corresponds to strength of prior reinforcement. Random and "No Prior" control TRNs serve as references in both panels. The right panel shows precision-recall of the KO-ChIP G.S. (9 TFs, 2375 edges) for TRNs built from the ATAC prior. **(B) Number of targets per TF in the Gold Standards.** Targets per TF are limited to the 3578 considered by the model. **(C) TF-specific TRN performance.** For each G.S., AUPRs were calculated for each TF individually. TF-specific performance of TRNs is quantified as the $\log_2$-foldchange between AUPR of the TRN model relative to random. Triangles indicated strength of prior reinforcement, and "m" and "TFA" denote methods of TFA estimation as in **(A)**. The TF-specific AUPR of input priors are displayed for reference.

Yosef et al. 2013). There are also 9 TF ChIP-seq experiments (Th17 48h time point), from which TF-gene interactions were derived based on ChIP peak proximity to gene locus (Ciofani et al. 2012). However, gold standards derived from either information source have some drawbacks. For example, differential analysis of TF KO expression profiles is imperfect, as cellular TRNs adapt to the original TF perturbation over time, leading to partial compensation by parologs (false negatives) and changes in expression of TFs and signaling molecules downstream of the original TF (false positives). The TF ChIP G.S. will also contain false positives (TF ChIP peaks are not necessarily functional) and false negatives (without access to 3D-chromatin conformation in Th17 cells, we limit to linear distances between TF ChIP peak and gene loci). Generating a gold standard from the intersection of edges supported by both TF KO and TF ChIP is likely an improvement over the individual G.S.'s. The resulting KO+ChIP G.S. should have a lower false-positive rate, but likely at the expense of more false-negatives. Given these caveats, we do not expect perfect recovery of the gold standards from inferred TRNs, but we do expect that the quality of inferred TRNs should scale with the precision-recall of KO+ChIP-supported edges.

For each G.S. information source, **Table S1A** summarizes the number of edges, TFs, and target genes as well as percent overlap with other prior information sources. **Figure 2B** shows the TF degree (number of gene targets) in each G.S. Because we have KO data for 25 TFs but KO+ChIP data for only 9 TFs, we include precision-recall analysis of the KO G.S. in addition to evaluating with the KO+ChIP G.S. Our KO-derived gold standard also enables testing of an initial, fundamental assumption: integration of high-quality TF DNA-binding data (from the TF ChIP experiments) will improve Th17 TRN inference. This is an important initial test of methods before moving to the noisier ATAC-seq prior. **Figure 2A** (left panel) displays the results of the analysis. The light-blue and pink curves describe performance without integration of prior information (based on gene expression alone) for BBSR-BIC and mLASSO-StARS, respectively. Integration of prior information outperforms the No Prior control in all cases for BBSR-BIC (darker blue curves) and all but one case for LASSO-StARS (red curves). Thus, TF ChIP data improves Th17 TRN inference. The cases explored include prior-based TFA estimation (with no, moderate, and high levels of prior reinforcement) and TF-mRNA-based TFA (with moderate and high prior reinforcement). For all but one case (prior-based TFA without reinforcement), mLASSO-StARS methods outperform the "equivalent" BBSR-BIC method. This trend also holds for the other prior information sources tested below (**Figures 2A,C, S3**). Importantly, we verify that levels of prior reinforcement were roughly equivalent for BBSR-BIC and mLASSO-StARS, plotting percent prior edges in the network as a function of network size (**Figure S2**). We also highlight that, for similar levels of prior reinforcement, prior-based TFA models outperform TF mRNA, and this trend also holds for the other priors tested.

We next tested inference with an ATAC prior (**Figure 2A** (central and right panels), **Figures S3**). Precision-recall of the KO G.S. enable direct comparison with the ChIP prior. Although the ATAC prior provides a boost over No Prior TRNs for both methods, the effect is less dramatic, likely reflecting increased levels of noise in the ATAC-seq prior. mLASSO-StARS methods with the ATAC prior show impressive precision-recall on the likely higher-quality G.S. (KO+ChIP, **Figure 2A** right panel). In contrast to results with a ChIP prior, increasing the strength of prior reinforcement from moderate to high yields no advantage for the noisier ATAC prior. This suggests ATAC-seq prior reinforcement should be limited to moderate rather than high (i.e., gene expression data should be relied on to select a small subset of the regulatory hypothesis from the ATAC-seq prior network).

To explore experimental designs that do not include context-specific ATAC-seq, we tested integration of other prior information sources on inference. Here, we show results for two very different publically available resources. The first is TF motif analysis of ENCODE DHS data from

25 mouse tissues, none of which include Th17 (Stergachis et al. 2014). The second is derived from the curated TRRUST database of human TF-gene interactions (Han et al. 2015). **Table S1A** provides high-level statistics on the resulting priors. While the ENCODE DHS prior includes ~1.5 million interactions between 546 TFs and ~17k genes (similar scale to the ATAC priors), the TRRUST prior is sparse: ~7k interactions between 582 TFs and ~2k genes. The TRRUST and ENCODE priors overlap less with the gold standards than context-specific priors, and this is reflected in lower precision-recall, relative to ChIP and ATAC priors (**Figures S3**). However, use of either the ENCODE or TRRUST prior improves performance relative to the "No Prior" control; this improvement is substantial for prior-based TFA models.

To evaluate method performance with fewer gene expression samples, we reduced the gene expression matrix from 254 to 50 randomly selected samples and evaluated performance on five independent subsamples (**Figures S4**). The smaller sample size had minor impact on precision-recall performance, especially for context-specific ChIP and ATAC priors, suggesting that these priors significantly guided TRN learning. Most conditions have one or more biological replicates, and this might partially explain the hardly-diminished performance using smaller gene expression datasets. These results bode well for extension of our methods to cellular contexts where gene expression data are less abundant.

We also evaluated methods at the resolution of individual TFs, to see how the different modeling decisions (e.g., choice of prior) affected target prediction for each TF (**Figure 2C**). There is nearly an order-of-magnitude difference in degree (number of targets) per TF in the KO+ChIP G.S. (**Figure 2B**), so this analysis additionally ensures that overall network results are not dominated by a few high-degree TFs. Overall, mLASSO-StARS also outperforms BBSR-BIC at TF-specific AUPR resolution (**Figure 2C**).

Closer inspection of trends reveals that the AUPR for many TFs was dependent on TFA estimation procedure (**Figure 2C**). TF mRNA can be a good estimate of protein TFA for TFs whose main source of regulation is transcriptional, while, for ubiquitously expressed TFs regulated by posttranslational modification, prior-based TFA would be expected to outperform TF mRNA estimates. Consistent with this intuition, models with prior-based TFA have higher AUPR for STAT3 (ChIP prior with KO G.S., ATAC prior with KO+ChIP and ChIP G.S.). In contrast, prior-based TFA was not best for recovery of RORC targets. With the ChIP prior (which includes RORγt TF ChIP), prior-based TFA AUPR was on par with TF mRNA AUPR. However, for the noisier ATAC-seq prior, prior-based TFA performed little better than random, while TF mRNA models (including the "No Prior" control) performed well (all G.S.'s). For ATAC-seq-based TRN inference, target prediction for some TFs (HIF1A, STAT3, NFE2L2) is better using prior-based TFA, while TF mRNA is better for some TFs (RORC, MAF, FOSL2) and roughly equivalent for others. Based on these results, we later generate "final" Th17 TRNs using a combination of prior-based and TF-mRNA TFA estimates.

Th17 TRN models predict out-of-sample gene expression patterns

We next evaluated whether the TRN models could predict out-of-sample gene expression patterns. In contrast to precision-recall (limited to the subset of TFs in the gold standard), gene expression prediction provides the opportunity to evaluate all interactions in the model. Thus, gene expression prediction is (1) an orthogonal means to quantitatively evaluate models and (2) especially important in poorly characterized cellular systems, for which gold standards do not exist and precision-recall is not possible.

Given the significant number of biological replicates in the Th17 gene expression data, we chose three out-of-sample prediction challenges for the TRN models. Each leave-out set of samples has

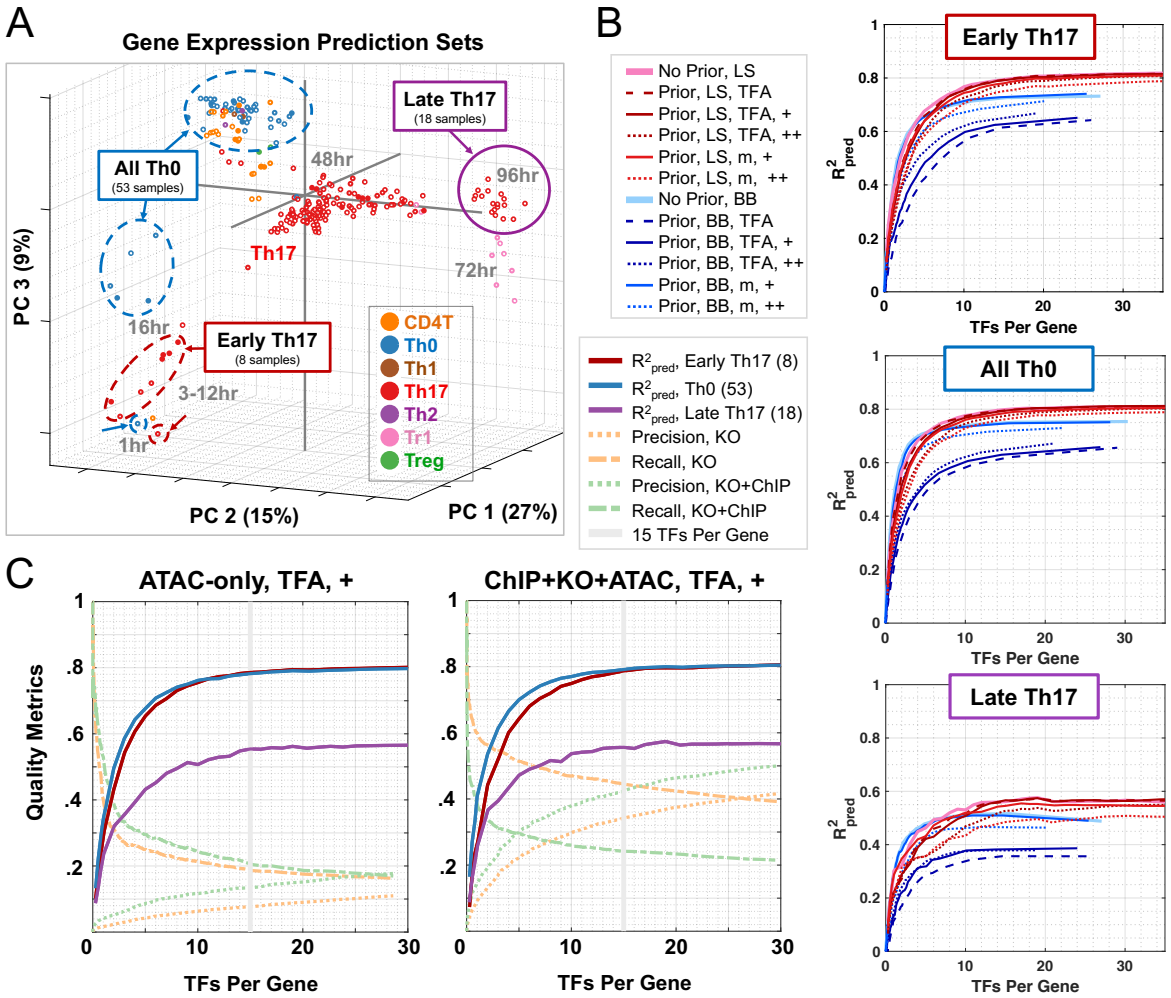

Figure 3. (A) Leave-out sets for gene-expression prediction. Model selection and parameter estimation of TF-gene interactions were performed in the absence of specified leave-out sets (All Th0, Late Th17, and Early Th17), circled in PCA space. **(B) Gene expression prediction by method.** R^2_{pred} for each leave-out set is plotted as a function of average number of TFs per gene. In the legend, LS = mLASSO-StARS, BB = BBSR-BIC, m = TF mRNA, TFA = prior-based TFA estimate, "+" corresponds to strength of prior reinforcement. **(C) Model quality metrics versus model size.** For two TRN models built with ATAC (left panel) or ChIP+KO+ATAC (right panel) priors (mLASSO-StARS, bias = .5, TFA = P^+X), the quality metrics (R^2_{pred} for each leave-out set, precision and recall) are plotted as a function of model size. The model size used for subsequent analysis is also highlighted.

distinct patterns of gene expression, highlighted in the PCA plot (**Figure 3A**). Specifically, we performed model selection and parameterization leaving out "Early Th17" (all Th17 time points between 1-16 hours, 8 samples), "All Th0" (Th0 samples for all time points and perturbations, 53 samples), or "Late Th17" (18 Th17 samples from 60-108h post-TCR stimulation).

Both BBSR-BIC and mLASSO-StARS methods provide confidence estimates for TF-gene interaction predictions; thus, for each method, we built models and tested prediction over a range of edge confidence cutoffs (each corresponding to a model size) (**Methods**). We quantified predictive performance using r-squared of prediction, R^2_{pred}, whose values range from one (perfect prediction) to $-\infty$; R^2_{pred}>0 indicates that the model has predictive benefit over the per-gene means of previously observed expression values (more detail in **Methods**). In **Figure 3B**, we compare predictive performance versus model size for TRNs built with an ATAC-seq prior as

well as "No Prior" controls for each leave-out set. Across all leave-out sets and methods, out-of-sample prediction improved most as models expanded from average size of 1 TF/gene to 5 TFs/gene. Predictive performance plateaued at ~10-15 TFs/gene, depending on the leave-out set. Most methods performed similarly well from 0-10 TFs/gene ("No Prior", all moderately- or un-reinforced mLASSO-StARS, moderately-reinforced, TF-mRNA BBSR-BIC methods), with the exception of BBSR-BIC models using prior-based TFA, whose prediction was substantially worse. For model sizes 10-15 TFs/genes, the mLASSO-StARS models had a performance gain over BBSR-BIC models. These results, together with the precision-recall results, support mLASSO-StARS methods over BBSR-BIC for mammalian TRN inference. In addition, moderately reinforced mLASSO-StARS models predicted slightly better than the other mLASSO-StARS models in terms of precision-recall or out-of-sample prediction; thus, we use mLASSO-StARS with moderate prior reinforcement to construct "final" Th17 TRNs.

Having performed these initial comparisons, we visualized all quality metrics (precision, recall, R^2_{pred}) versus model size to select a model-size cutoff. **Figure 3C** shows results for moderately reinforced prior-based TFA models with an ATAC ("ATAC-only") or ChIP+KO+ATAC prior, while results for "No Prior" control and other conditions can be found in **Figure S5**. The ChIP+KO+ATAC prior represents our best (combined) source of prior information and will be later used to generate a final Th17 TRN (see ChIP/A(Th17)+KO+A(Th) prior description in **Methods**). Once model sizes reach ~15 TFs/gene, predictive performance plateaus, suggesting ~15 TFs/gene as a potential cutoff for edge inclusion in the network (see more detailed discussion in **Methods**). Of note, our gene expression matrix includes a variety of T Cell polarization conditions, time points and other perturbations. Thus, at a model cutoff of 15 TFs/gene, only a subset of TFs in a gene model are expected to drive gene expression in any single condition. For ATAC-only TRNs, a cutoff ~15 TFs/genes corresponds to precision of ~20% and recall of 15% and 10% for KO+ChIP and KO G.S.'s, respectively. For the ChIP+KO+ATAC TRN, the ChIP+KO interactions are included in the prior, and so precision and recall does not measure predictive performance but rather indicates how many KO+ChIP interactions are contained at each model size. At a cutoff of ~15 TFs/gene, ~45% of model edges are KO+ChIP supported, while ~25% are supported by KO ("precision" curves), while ~40% or 35% of the KO+ChIP or KO prior interactions make it into the model ("recall" curves). This pruning behavior suggests that multivariate modeling of gene expression patterns could filter false-positives TF-gene interactions from the KO+ChIP prior. Equally important, 55% of the Th17 TRN edges were not present in the initial KO+ChIP network but are supported by gene expression and ATAC-seq data. Thus, integration of new data expands our model of Th17 gene regulators and regulation.

Source of prior, TFA method and prior reinforcement dramatically alter Th17 TRNs

Precision-recall analysis, limited to the 9-25 TFs in the gold standards, revealed some differences among the mLASSO-StARS models, while all models had roughly similar out-of-sample gene-expression prediction. Having performed these initial quality assessments, we explore the Th17 TRNs globally.

Standardizing network sizes to 53k TF-gene interactions (~15 TFs/gene), we calculated the percent edge overlap among TRNs built from ATAC, ChIP, KO, ENCODE DHS or TRRUST priors as well as two combined priors, expected to be improvements relative to their individual constituents. The first combined prior "ChIP/A(Th17)" uses TF ChIP data when available and TF motif analysis of 48h Th17 ATAC-seq data otherwise. The second prior "ChIP/A(Th17)+KO+A(Th)" combines the first with KO and ATAC-seq data from the other time points and conditions (**Methods**). For each prior, we considered five modeling modes: prior-based TFA estimates ("TFA") with no, moderate or strong prior reinforcement and TF mRNA TFA estimates ("m") with moderate and strong prior reinforcement. We also included the No Prior TRN

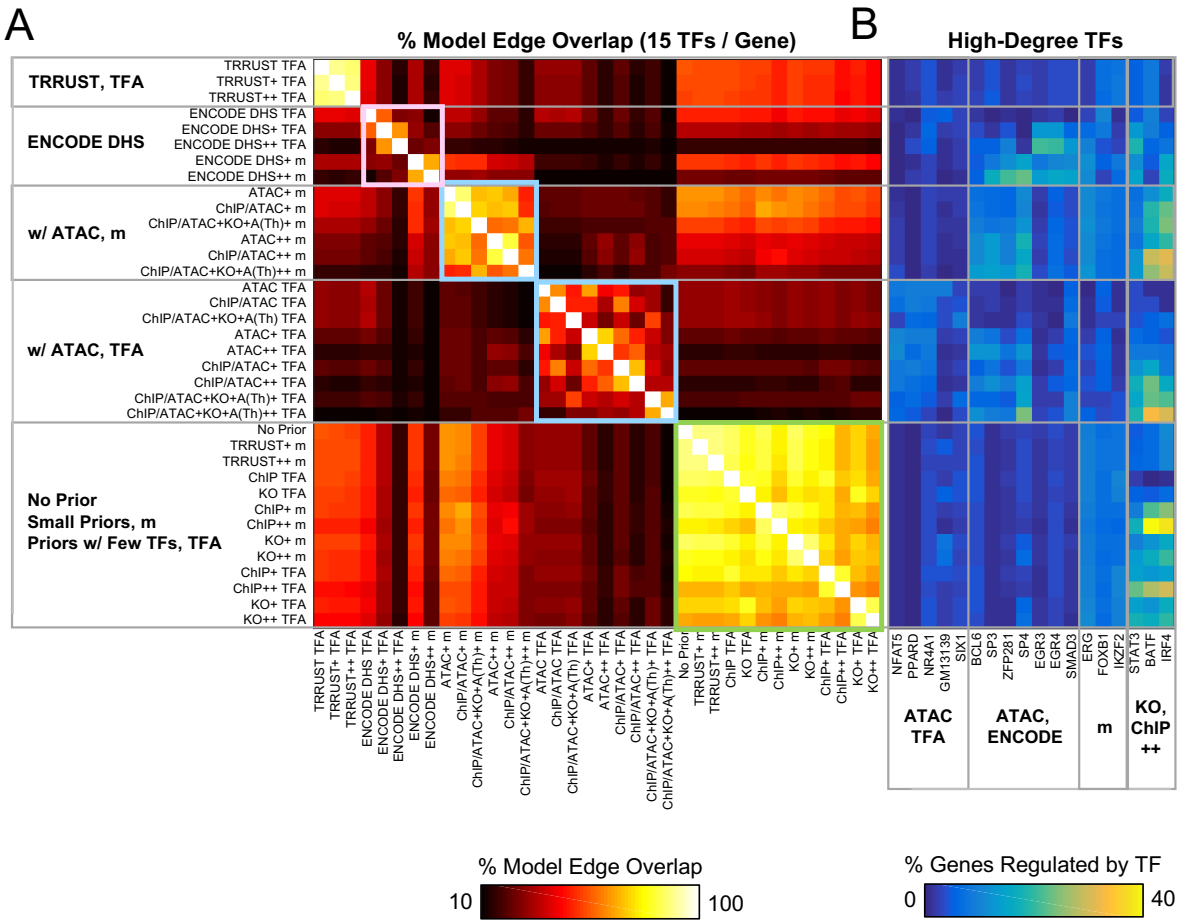

Figure 4. (A) Edge overlaps of TRNs. TRNs were built with ATAC, ChIP, KO, ENCODE DHS, TRRUST, ChIP/ATAC or ChIP+ATAC+KO priors, with TFA based on target genes ("TFA") or TF mRNA ("m") and varying strengths of prior reinforcement (+ = moderate, ++ = strong, or none). The "No Prior" TRN was included as a control. Each TRN models size was limited to a mean of 15 TFs per gene and hierarchically clustered using 1-overlap distance and ward linkage. **(B) Network coverage by high-degree TFs.** A TF was included in the "high-degree" TF set, if it was within the top-two highest degree nodes in one or more of the TRNs. The percentage of genes regulated by TF is based on the 3578 target genes.

as a control. The results described below, for network sizes ~15 TFs/gene hold for other network-size cutoffs (5, 10, and 20 TFs/gene, data not shown). The percentage of shared edges between TRNs ranged from 83% to 10%. Clustering revealed several trends (**Figure 4A**). Over a third of the TRNs cluster together with high overlaps (60-88%), highlighted by a green-outlined box in **Figure 4A**. We refer to this cluster as "gene-expression-driven", because it contains the "No Prior" TRN and other TRNs where TF-gene interactions are mainly based on TF-mRNA TFA estimates. The cluster contains all TF-mRNA models for the smaller priors (ChIP, KO, TRRUST) as well as some models built with prior-based TFA, but only for priors with few TFs (ChIP and KO) and for which TF mRNA must be used for the remaining TFs. Consistent with this explanation, the TRRUST TRNs built with prior-based TFA cluster separately. Although the TRRUST prior has fewer edges (~1500) than either the ChIP (~13k) or KO (~8k) prior, its edges are distributed among 187 of the 715 potential TF regulators considered (**Table S1B**), and, thus, TRRUST-based TFA greatly alters the resulting TRN.

Three priors contain ATAC-seq data (ATAC, ChIP/A(Th17), and ChIP+KO+ATAC); these priors are large (ranging from ~120-200k interactions for 346 TFs), so large that not even moderately

reinforced, TF mRNA TRNs cluster with the "gene-expression driven" TRNs. The ATAC-containing TRNs form two clusters (marked by light-blue boxes in **Figure 4A**) that separate based on TFA estimation method. As expected, the "ATAC TF mRNA" cluster is more coherent (overlaps ranging from 47-87%), while overlaps in the "ATAC prior-based TFA" cluster range from 17-70%. The ATAC TRNs built on TF mRNA are most similar to the "gene-expression driven cluster", but with only ~50-65% edge overlap for moderate-strength prior reinforcement. At high prior reinforcement, overlap drops to ~35-45%. The "ATAC prior-based TFA" cluster is as dissimilar to the "ATAC TF mRNA" cluster as it is to the "gene-expression driven cluster" (overlaps with both clusters ranging from 10-30%). Thus, as also suggested by the TRRUST cluster, prior-based TFA involving a large number of TFs, dramatically changes the resulting TRN. The ENCODE prior produced the most dissimilar TRNs, both when considering ENCODE TRNs relative (1) to each other and (2) to TRNs based on other priors. This result makes sense, given the unique origin and large size of the ENCODE prior (~200k interactions for 238 TFs).

To further our intuition about these distinct TRNs, we visualized the percentages of genes regulated by TFs with high degree across TRNs (**Figure 4B**). The high-degree TFs can be roughly divided into four clusters. Some top-degree TFs are mainly driven by gene expression (ERG, FOXB1, IKZF2, denoted by an "m" in **Figure 4B**) and regulate ~7-9% of the 3578 network genes not only in the "gene-expression TRNs" cluster but also TF mRNA-based ENCODE and ATAC-containing networks. Moderate to strong reinforcement of ChIP and KO priors lead to high degree for key Th17 TFs: BATF, IRF4, and STAT3. In the No Prior TRN, the percent of genes regulated by these TFs hovers between 4-10%, while, for strongly enforced ChIP TRNs, the percent genes regulated climbs to 17-40%. Strongly reinforced priors composed of ChIP data in combination with ATAC and/or KO data also leads to high-degree for this TF cluster, ranging from control of 15-32% of genes. Priors with few TFs (e.g., KO and ChIP) lead to models that contain >75% edge overlap with the No Prior TRN across models, but this high overlap can be deceptive. Even for small priors, the modeling decisions can significantly alter the TF degree distribution.

The differences in network overlap and high-degree TFs between methods might reflect different strengths of each method and prior information source. Although overall-network performance in terms of precision-recall is roughly similar between methods and priors, the TF-specific AUPRs were heterogeneous by method (**Figure 2C**). Thus, each method enriched for unique sets of TF-specific gold-standard interactions among its high-confidence edges, and the ensemble of predictions from each model and method might provide a fuller picture of the underlying TRN.

"Core" Th17 TRNs contain literature-supported TF-gene interactions

Our primary objective is to assess the feasibility of high-quality TRN inference from gene expression and ATAC-seq data. Therefore, it is important to examine the Th17 TRNs at high resolution. Here, we focus analysis on TRN predictions for 18 "core" Th17 TFs and genes, readily familiar to Th17 biologists (**Figure 5**). Individual interactions between TFs and genes in this "core" have been the focus of many studies (Christie et Zhu 2014; Li et al. 2014), which we leverage to evaluate the ATAC-based Th17 core TRNs: moderately reinforced TF-mRNA- and prior-based TFA TRNs individually and combined into a final ATAC-only TRN (as described in **Methods**). To visually compare these networks, we developed jp_gene_viz, a Jupyter-notebook TRN visualization tool. In addition to the ATAC TRNs (**Figure 5A**), we include the KO-ChIP G.S. (without inference) and the No Prior TRN as references (**Figure 5B, 5C**). For more detailed exploration, all 36 LASSO-StARS Th17 TRNs (from **Figure 4**), gold standards, and final, combined TRNs are available in a Jupyter-notebook binder: https://mybinder.org/v2/gh/simonsfoundation/Th17_TRN_Networks/master.

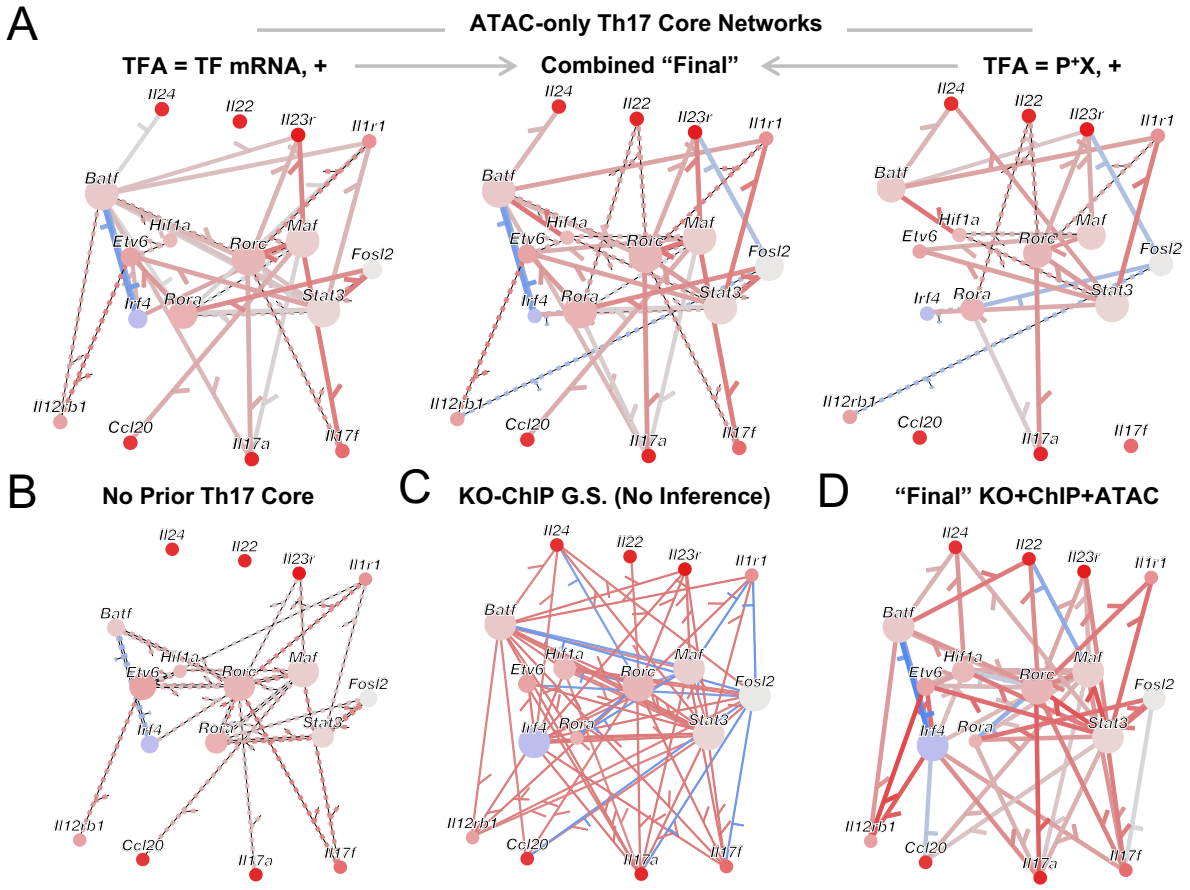

Figure 5. (A-D) Th17 core TRN models. "Core" Th17 genes and TFs were selected from the literature for visual comparison with jp_gene_viz software. Network size was limited to an average of 15 TFs per gene for Inferelator networks using: **(A)** ATAC prior, **(B)** no prior, and **(D)** ChIP+ATAC+KO prior. The edges in Inferelator TRNs are colored according to partial correlation (red positive, blue negative) and weighted relative to edge stability. **(C)** represents the full KO-ChIP G.S. from Ciofani et al., where edge sign is based on differential gene expression analysis between TF KO and control and edge weight is constant. Nodes are colored according to gene expression at 48h in Th17, z-scored relative to Th17 and other T Helper cell time points (red/blue = increased/decreased expression).

From the literature and the KO-ChIP G.S. (**Figure 5C**), there is support for edges between RORC and several key Th17 cytokines and receptors: *Il17a*, *Il17f*, *Il22*, *Il1r1* and *Il23r*. Two of these interactions (*Il17a*, *Il23r*) were present in the ATAC-seq prior and both TF-mRNA- and prior-based TFA ATAC TRNs (**Figure 2A**, left and right). Four of the interactions (*Il17a*, *Il17f*, *Il1r1* and *Il23r*) are recovered on the basis of gene expression modeling alone, as they were recovered by the No Prior TRN and also the TF-mRNA ATAC TRN. Interestingly, the prior-based TFA ATAC TRN (right panel **Figure 5A**) recovers a data-driven edge between RORC and *Il22*. By combining predictions from both ATAC TRNs (see **Methods**), all five RORC targets are recovered (central panel, **Figure 5A**). As expected, inclusion of RORC ChIP and/or KO in the prior leads to recovery of all five TF targets as well (e.g., **Figure 5D**).

STAT3 is required for Th17 differentiation, playing a crucial role in the activation of *Rorc*. There is support for this interaction not only from context-specific ATAC-seq prior but also the ENCODE DHS prior. Consistent with the poor quality of mRNA-based STAT3 predictions (**Figure 2C**), this interaction is not present in the No Prior TRN, but is recovered in ATAC and ENCODE TRNs,

even the TF mRNA models, where prior reinforcement likely overcomes weak correlation between TF mRNA and protein activity level (**Figures 5A-C, S6**).

MAF is another key regulator of Th17 cytokine and receptor expression, with KO-ChIP G.S. edge support for *Il17a*, *Il17f*, *Il23r*, and *Il1rb*. There is ATAC-seq support for MAF regulation of *Il17a*, *Il17f*, *Il23r*. The prior-supported targets are recovered by the TF-mRNA-based ATAC models, but only the *Il23r* interaction is present in prior-based TFA models (**Figure 5A**). Similar to RORC, this suggests the signal between *Maf* mRNA and target genes might have been better than the ATAC-inferred MAF TFA. Prior reinforcement also played a role, as only two of the four interactions are present in the No Prior TRN (**Figure 5B**). In the absence of context-specific prior information and a strong signal from the gene expression model, only a single edge (*Il23r*) was recovered by one of the ENCODE models (**Figure S6**).

These results highlight the potential for TRN inference in new settings, where integration of chromatin accessibility and gene expression is more feasible than sequential TF ChIP and KO experiments. They also suggest that there is value to building models from both TFA methods, even though, prior-based TFA dominated precision-recall performance overall (**Figures 2A, S3, S4**). For construction of the "final" Th17 TRN, we combine models based on both TFA methods, choosing only moderate levels of prior reinforcement to accommodate useful but noisy ATAC-seq data in our final combined prior (**Figure 5D**). The literature-curated core of our final Th17 TRN contains the RORC and MAF cytokine and receptor interactions highlighted from the literature, as well as the established connection between STAT3 and *Rorc*.

ATAC-derived Th17 TRNs contain known and novel Th17 transcription factors

Having verified that the Th17 TRNs contain core Th17 TF-gene interactions from the literature, we develop a global, unbiased analysis of the final ChIP+ATAC+KO TRN to identify new network predictions coordinating relevant Th17 biology. In addition, we extend our analysis to a final TRN an ATAC-only prior, to simulate mammalian TRN inference in less well-studied systems, where KO and/or ChIP data might not be available.

Our analysis begins with a high-level view of TRN interactions (**Figures S7, S8**), where we rank TFs by degree, dividing interactions on the basis of sign and whether they were in the prior. Overall, prior-supported edges make up 63% and 43% of the ~53k TF-gene interactions in ChIP+KO+ATAC and ATAC-only TRNs, respectively. Of the 715 potential TF regulators considered for final models, nearly all (~95% have targets in the final TRNs, with positive interactions outnumbering negative nearly two-fold (1.8:1 (ChIP+KO+ATAC) or 1.9:1 (ATAC-only)). TF degree varies dramatically (**Figures S7, S8**). Whereas the ATAC-only network democratizes TF degree distribution (no TF has > 500 targets), the addition of ChIP and KO leads to very high degree for several TFs in the ChIP+KO+ATAC TRN (>500 targets for IRF4, BATF, MAF, SP4, FOSL2, and STAT3). While there is a bias for TFs in the prior to have higher degree, several TFs without prior support have >100 targets in the final networks (4 TFs for the ChIP+KO+ATAC TRN and 14 TFs for the ATAC-only TRN). Thus, prior information is not weighted so strongly as to preclude inclusion of the many TFs without known motifs. This aspect is important for discovery and also holds for TFs with edges in the prior, too. While only 16% or 7% of input prior edges remain in the TRN (edges without support in the gene expression model drop out, likely reducing false positives), 27% or 46% of learned regulatory interactions for TFs with prior information were not originally in the prior (percentages for KO+ChIP+ATAC or ATAC-only TRNs, respectively). This is also a desirable property, given the presence of false negatives in our priors. For example, motif analysis of the RORC ChIP data revealed that only ~1/3 of RORC peaks contained a RORC motif (at motif occurrence cutoff P_{raw}=1E-4). Although RORC can bind DNA directly, the RORC ChIP data suggests that RORC might also bind DNA indirectly (e.g., via

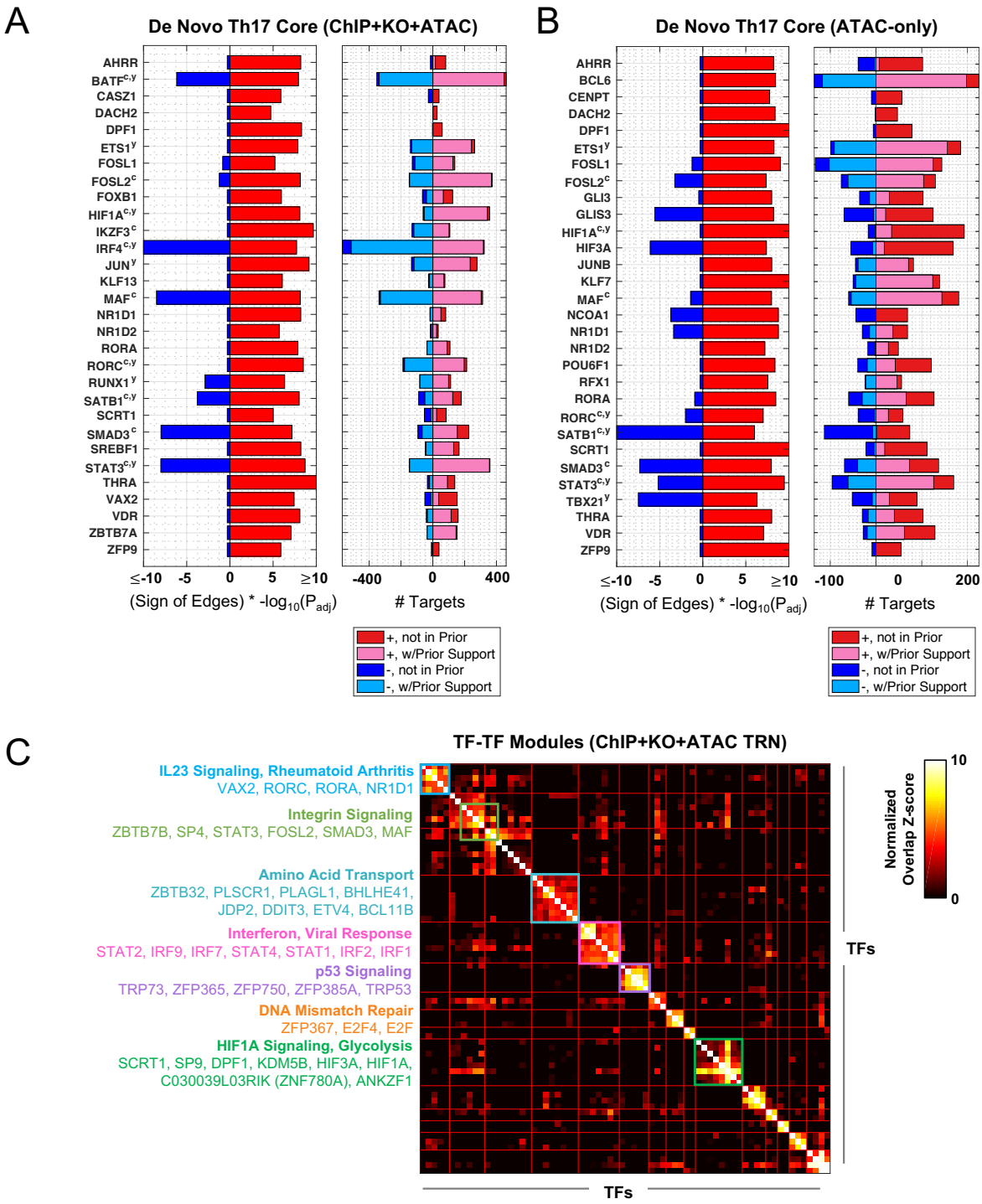

Figure 6. De Novo Th17 Core TFs in the ChIP+KO+ATAC TRN (A) and ATAC-only TRN (B). TFs were included as De Novo Th17 Core TFs if, at a FDR = 10%, (1) positive gene targets were enriched for up-regulated Th17 genes or cytokines and receptors or (2) negative gene targets were enriched for down-regulated Th17 genes or cytokines and receptors. Left panel indicates enrichment in these sets. Right panel denotes number, sign and prior support of TF target edges. Superscripts c and y indicate TF Th17 association from (Ciofani et al. 2012), (Yosef et al., 2013), respectively. **(C) Top 15 TF-TF Modules for ChIP+KO+ATAC TRN**. TFs were clustered into modules on the basis of shared positive target genes between TFs (see **Methods**). The Top 15 TF-TF modules are displayed. Gene-set enrichment was used to annotate clusters, and TF members are listed. Asterisk indicates TF-TF modules that contain Th17-promoting core TFs.

TF complexes). Such indirect binding would be difficult to detect by ATAC-seq motif analysis alone, but their gene targets can be recovered via gene expression modeling.

Many of the TFs with highest degree are shared between the ChIP+KO+ATAC and ATAC-only TRNs (e.g., IRF4, BATF, SP4, RXRA, STAT3,... **Figure S7, S8**). While the so-called "master regulator" of Th17, RORC, has the 11th-highest degree in the ChIP+KO+ATAC TRN, RORC is absent from the top-100 highest degree TFs of the ATAC-only TRN. Superficially this appears to be a negative result, but the TRNs explain expression patterns for genes that are both up-regulated *and* down-regulated in Th17. For that reason, degree is not the best way to define a Th17 core de novo, and we developed an unbiased approach to identify TFs that specifically promote Th17 expression patterns. TFs were included in our core if they met one of two criteria: (1) The TF promotes Th17 gene expression through activation of Th17 genes (i.e., the TF's positive edges are enriched in up-regulated Th17 genes at an FDR = 10%) or (2) The TF promotes Th17 expression through repression of non-Th17 genes (i.e., TF's negative edges are enriched in down-regulated Th17 genes at an FDR = 10%). This analysis yielded strikingly similar de novo Th17 core TFs for both ChIP+KO+ATAC and ATAC-only TRNs. **Figure 6A,B** highlights the Top-30 core TFs with several recognizable Th17-specific TFs from the literature (RORA, RORC, STAT3, MAF) in both networks.

Given the size of the Th17 TRNs, we designed a TF-TF module analysis to define clusters of TFs with significant overlap in target genes (see **Methods**, **Figures 6C, S9, S11**). We then applied a comprehensive gene-set enrichment analysis to predict functional roles for the TF-TF clusters, looking for consensus among pathway enrichments from five databases (Gene Ontology, Pathway Commons, KEGG, WikiPathways and signatures from MSigDB (Consortium 2004; Cerami et al. 2010; Kanehisa et Goto 2000; Pico et al. 2008; Liberzon et al. 2011)) (**Figures S10, S12**). Most TF-TF clusters were conserved between ChIP+KO+ATAC and ATAC-only networks, and, within clusters, TFs shared features. For example, several clusters contained TFs defined in the de novo Th17 cores. RORC was a member of a Th17-promoting TF-TF Module including RORA, NR1D1, VAX2 (highlighted with a light-blue square, **Figures 6C, S9-S12**); functional annotations for this cluster include "IL23 Signaling", "Rheumatoid Arthritis" and are consistent with prior knowledge. Th17-promoting genes HIF1A, HIF3A, DPF1, SP9, and SCRT1 cluster with 5-6 other TFs (green square), and enrichments for this cluster include "hypoxia", "HIF1A transcription factor network", and "glycolysis".

Other clusters contained TFs that promote the expression of genes repressed at 48h in Th17 cells. One such cluster contained Th1 TFs (IRF1, STAT1 and STAT2), many other interferon response factors (IRF2, IRF7, IRF9) as well as STAT2 and STAT4 (hot-pink box, **Figure 6C**). As expected, this "interferon cluster" has enrichments for "Response to Interferon Gamma", "Type 1 Interferon Pathway", "Response to Virus". Interestingly, although TF gene expression for this cluster is highest in Th1 relative to other T Cell populations at 48hrs, gene expression is at its highest at the 1hr Th17 time point, suggesting an interferon-like response for Th17 cells very early in the Th17 polarization time course (**Figures S9, S11**). This result is consistent with predictions from another Th17 TRN (Yosef et al. 2013), in which authors also predict roles for IRF1, IRF2, IRF9, STAT1, STAT2 within the first four hours of Th17 polarization. Both findings are consistent with potential plasticity, observed *in vivo*, in Th17 cell programs that are homeostatic or pathogenic, with expression of Th1-like features in the latter.

Gene-set enrichment provides functional predictions for other TF-TF Modules, including: (1) "Amino Acid Transport" (ZBTB32, PLSCR1, PLAGL1, BHLHE41, JDP2, DDIT3, ETV4, BCL11B), (2) "Integrin Signaling" (ZBTB7B, SP4, STAT3, FOSL2, SMAD3, MAF), (3) "DNA Mismatch

Repair" (ZFP367, E2F4, E2F), (4) "p53 Signaling" (TRP73, ZFP365, ZFP750, ZFP385A, TRP53), and others (**Figures 6C, S10, S12**). These predictions provide confirmation of TRN quality, as many modules have predicted function in processes for which individual TFs are already implicated (e.g., HIF1A and HIF3A appear in the HIF1A/hypoxia module, TRP53 and TRP73 appear in the p53-signaling module). In addition, TF-TF modules and functional annotations are largely conserved between KO+ChIP+ATAC and ATAC-only TRNs, the latter prior network being much more economically feasible than the first. More fundamentally, these predictions suggest how altering sets of TFs might influence Th17 pathways and responses.

Discussion

Th17 cells protect mucosa from bacteria and fungi but can also drive autoimmune and inflammatory disease (Khader et al. 2009; Littman et Rudensky 2010; Stadhouders et al. 2017). These diverse roles require coordination of thousands of genes. TF-specific knowledge of gene expression regulation is of great interest, providing a map for immuno-engineering Th17 behavior in disease. Researchers in academia and industry have used our first genome-scale Th17 transcriptional regulatory network (Ciofani et al. 2012) to develop hypotheses in the context of autoimmunity (Patel et Kuchroo 2015; Yang et al. 2014; Isono et al. 2014). Here, we provide an important update to our knowledge of Th17 transcriptional regulation, enabled by technical advances in genomic measurement and computational advances in TRN inference. KO data for 20 TFs and TF ChIP data for 9 TFs were central to the original Th17 TRN, providing excellent coverage of TF-gene targets for TFs in that set and a resource for the Th17 community. However, technical limitations and cost precluded application of these tools to the hundreds of TFs expressed over the course of Th17 differentiation, all of which could play important roles in Th17 gene expression regulation.

In combination with large-scale efforts to learn TF DNA-binding motifs (Weirauch et al. 2014; Jolma et al. 2013; Najafabadi et al. 2015; Badis et al. 2009), the advent of ATAC-seq represents an opportunity to overcome limitations of sequential TF ChIP experiments, expanding the number of TFs with chromatin binding profiles by over an order of magnitude (**Table S1B**). In addition, although TF KO and ChIP data were pragmatically limited to 48h Th17 conditions, standard ATAC-seq protocols require two orders of magnitude fewer cells than TF ChIP and can be applied directly to cells from humans or animal models. Here, we were able to obtain (indirect) TF binding profiles from multiple Th17 differentiation time points. Yet TF-binding profiles derived from motif analysis of ATAC-seq are noisy, containing false positives and false negatives. In addition, only ~40% of mouse or human TFs have known motifs, so motif analysis of ATAC-seq data excludes more than half of potential TF regulators. Here, we provide a single, integrated method to infer regulatory roles for TFs genome-wide. At its core, gene expression is modeled as a function of TF activities, where prior information (e.g., from ATAC-seq) can be used to (1) improve TF activity estimates (and therefore target prediction) for some TFs and/or (2) favor TF-gene interactions that also have support from other sources (e.g., ATAC-seq, KO). We rigorously test the performance of our method in terms of precision-recall of gold-standard interactions from TF KO and ChIP and verify TRN model quality in terms of gene expression prediction. Our methods have two very desirable features (1) they prune initial noisy prior networks (by over an order of magnitude in this study) while (2) also learning new TF-gene interactions for TFs with and without prior information.

Our final Th17 TRN is built integrating our best knowledge for model-building (combining knock-out of key TFs, ChIP-seq of multiple Th17-relevant TFs, and ATAC-seq with a rich gene expression dataset). While the original "core" Th17 network was limited to TFs with ChIP and KO data, our de novo Th17 core includes the original core (RORC, STAT3, BATF, IRF4, MAF) and

dozens of additional TFs. We performed extensive gene-set enrichment to determine functional roles and potential phenotypes for novel core TFs. We detect TF-TF modules and make multivariate predictions for gene pathways regulated by multiple TFs. Although HIF1A was detected as an important TF regulating hypoxia-related genes in the original Th17 TRNs, here we strongly implicate several other factors (HIF3A, DPF1, SCRT1) in regulation of hypoxia genes. In addition, we uncover multivariate regulation of amino acid transport, interferon, and many other pathways. This work provides an important update to our knowledge of transcriptional regulation in Th17 cells.

Of perhaps greater importance, the TRN experimental design and computational methods proposed are general, designed for regimes where prior knowledge of transcriptional regulators and/or sample material is scarce (e.g., cells directly from humans and animal models). Given the rigorous testing and case study presented here, we have high expectations for their successful application in other systems. Indeed, de novo Th17 cores and TF-TF modules are remarkably similar between TRNs derived from all prior information (KO+ChIP+ATAC) and ATAC-seq data only. Our methods are also general in that prior information need not be limited to motif analysis of ATAC-seq data. TF binding patterns can also be inferred from TF ChIP, histone-mark ChIP, FAIRE-seq, and DHS. We also integrate prior information from other sources (differential analysis of TF KO data and the curated TRRUST database). Thus, our methods are widely applicable.

This work also highlights avenues for future improvement of TRN inference methods. We tested two methods for TF activity estimation: (1) TF mRNA levels and (2) based on prior knowledge of TF target gene expression. Although prior-based TFA improved TRN inference in B. subtilis and yeast (Arrieta-Ortiz et al. 2015; Tchourine et al. 2017), neither method consistently outperformed the other in this study. As a result, final TRNs were built using both TFA methods. There are multiple dimensions along which TFA estimation could be improved. The simplicity of the linear framework proposed for prior-based estimation has limitations in the context of complex mammalian transcriptional regulation, and a more sophisticated mathematical model for TFA estimation could be of value. TFA estimation would also improve from better prediction of TF binding events. Here, we limited our approach to a simple TF motif analysis of accessible chromatin, yet several more sophisticated methods exist and merit testing (Pique-Regi et al. 2011; Sherwood et al. 2014; Chen et al. 2017). Another limitation of our method is the crude way in which putative TF binding events are mapped to gene loci. In our analysis, 3D distance between potential regulatory regions and gene loci is approximated by linear distance, a short-coming that chromatin capture data (e.g., Hi-C (Lieberman-Aiden et al. 2009) and other 3D-chromatin techniques (Zhang et al. 2012; Beagrie et al. 2017) would mitigate. To this end and others, the Th17 genomics dataset (Ciofani et al. 2012), augmented by our new ATAC-seq and RNA-seq experiments, provides a fertile testing ground for the development of future TRN inference methods and innovation.

Methods

ATAC-seq

CD4+ T Cells were sorted and polarized according to (Ciofani et al. 2012), and ATAC-seq samples were prepared as described previously (Buenrostro et al. 2013). Paired-end 50bp sequences were generated from samples on an Illumina HiSeq2500. Sequences were mapped to the murine genome (mm10) with bowtie2 (2.2.3), filtered based on mapping score (MAPQ > 30, Samtools (0.1.19)), and duplicates removed (Picard). For each sample individually, we ran Peakdeck (McCarthy et O'Callaghan 2014) (parameters –bin 75, -STEP 25, -back 10000, -npBack100000) and filtered peaks with a P_{raw}<1E-4. To enable quantitative comparison of accessibility across samples, we generated a reference set of accessible regions, taking the union

(Bedtools) of peaks detected in individual samples. The reference set of ATAC-seq peaks contained 63,049 potential regulatory loci, ranging from 75 base pairs (bp) to 3725 bp (median length 275 bp). Reads per reference peak were counted with HTSeq-count. ATAC-seq data was robustly normalized using DESeq2 (Love et al. 2014) counts for PCA.

RNA-seq

Data Sources: RNA-seq data from Littman-lab publications was downloaded from GEO: GSE40918 (156 samples), GSE70108 (4 samples), and GSE92992 (8 samples). The 18 new samples composing "late Th17" time points were generated as follows: Naïve CD4 T cells were primed on an anti-CD3/CD28 coated plate (without any additional cytokine) for 12-14hrs (overnight) and then polarized the cells with one of two cytokine cocktails: (1) Th17N: TGF-b (0.3ng/ml) + IL6 (20ng/ml) and (2) Th17P: IL6 (20ng/ml) + IL1b (20ng/ml) + IL23 (20ng/ml). Cells were harvested for RNA-seq at 60 and 108hrs post-TCR stimulation. Cells were lysed and snap-frozen in trizol then thawed for chloroform extraction, and 70% ethanol was mixed in a 1:1 ratio with the aqueous phase. Samples were then loaded onto a Qiagen RNEasy column according to manufacturer's instructions. A Ribo-Zero kit was used to deplete rRNA; libraries were prepared and sequenced on an Illumina HiSeq2500. The remaining CD4+T cell samples were (1) naïve or polarized and (2) processed as described in (Ciofani et al. 2012).

Batch Effects: The final gene expression matrix contained samples generated over an eight-year period, from several genomics cores, and with a combination of 36-bp and 50-bp single-end reads as well as 50-bp paired-end reads. There was no common set of control samples across batches, and so we developed an alternative to popular batch-correction methods ComBat (Chen et al. 2011). We first attempted to control for differences in sequencing, mapping only a single-end from paired-end sequencing samples with STAR aligner (Dobin et al. 2013). In addition, we controlled for differences in mappability between 36-bp and 50-bp reads by calculating effective gene lengths. Specifically, we calculated the number of possible 36-mers or 50-mers that HTSeq-count would map per gene (using HTSeq-count parameters: --stranded=no --mode=union and 36-mer and 50-mer mappability tracks (Derrien et al. 2012). We then computed a factor (effective 36-mer gene length to effective 50-mer gene length) to convert 50-mer gene quantification by HTSeq-count to 36-mer gene quantification scale. From there, we used DESeq2 (Love et al. 2014) to normalize data. We used PCA to ensure that major trends did not correlate with (1) sequencing, (2) data of experiment, or (3) library size. Instead, they correlated well with the T Cell time course and polarization conditions (**Figures 1A, S1**).

Transcriptional regulatory network inference.

Selection of Target Genes: For testing and benchmarking of methods, we attempted to build gene expression models for 3578 genes, the union of (1) genes differentially expressed between Th17 and Th0 at 48h (FDR = 10%, $\log_2|FC| > \log_2(1.5)$) and (2) the 2100 genes in the original Th17 TRN (Ciofani et al. 2012). This list is available in **Table S2**.

Selection of Potential Regulators: We initially generated a custom list of potential mouse protein TFs, combining (1) mouse and human TFs from TFClass (Wingender et al. 2014) and genes with the GO annotation "Transcription Factor Activity". (Human TFs were mapped to mouse using the MGI database.) From our list of potential mouse protein TFs (2093 genes), we generated a final list of 869 potential TF regulators, limited to those TFs with differential gene expression in at least one pairwise comparison between Th17, Th0, Th1, Treg, or Th2 at 48h (FDR = 10%, $\log_2|FC| > \log_2(1.5)$). This initial list of 869 TFs was used for all analyses comparing mLASSO-StARS and BBSR-BIC (**Figures 2, 3B, S2, S3**). However, given very recent efforts in TF annotation (Lambert et al. 2018), we generated a new list of mouse TFs for all subsequent TRN analyses. Specifically, *Lambert et al*. began with a list of candidate human TFs and manually curated whether available

data supported a TF assignment, yielding both lists of likely TFs and likely nonTFs. We again used MGI to convert both lists to mouse. To gain mouse TFs without human orthologs, we integrated with mouse TFs from AnimalTFDB (Zhang et al. 2015), but removed any mouse TFs (70) that were unlikely to be TFs (Lambert et al. 2018). Our final mouse TF list contained 1577 TFs, 715 of which served as potential regulators (differentially expressed as described above). Both lists of candidate TFs are available in **Table S3**. Note that **Tables S1A** and **S1B** correspond to the initial and updated TF lists, respectively.

Generation of Prior Matrices:
ATAC-seq: ATAC-seq peaks were detected (as described above) and associated with putative transcription-factor (TF) binding events and target genes to generate a "prior" network, $P \in \mathbb{R}^{|genes| \times |TFs|}$, of TF-gene interactions. We used a compendium of human and mouse TF motifs. Human and/or mouse transcription factor binding motifs (PWMs) were downloaded from the CisBP motif collection version 1.02 (http://cisbp.ccbr.utoronto.ca) (Weirauch et al. 2014) and the ENCODE motif collection (http://compbio.mit.edu/encode-motifs) (Kheradpour et Kellis 2014). Transfac version 2014.2 motifs (Wingender 2008) referenced from the human CisBP collection were converted from transfac frequency matrix format to .meme format with the MEME Suite tool transfac2meme version 4.10.1. To add human ENCODE motifs to the collection, redundant entries for each TF were first removed from the ENCODE set if the PWM was highly correlated ($R^2 > 0.95$) with any CisBP entry for that TF. Human TF in that combined ENCODE and CisBP set were then mapped to mouse orthologs. We scanned peaks for individual motif occurrences with FIMO (Grant et al. 2011) (parameters --thresh .00001, --max-stored-scores 500000, and a first-order-Markov background model). We found inclusion of human TF orthologs from the ENCODE motif collection slightly increased precision-recall relative to mouse CisBP alone (data not shown). TF motif occurrences with raw p-value < 1E-5 were included in downstream analysis. Putative binding events were associated with a target gene, if the peak fell within +/-10kb of gene body. We tested several peak-gene association rules based on distance from gene body or TSS, and TRN inference was robust to that choice (data not shown). We generated two ATAC-seq priors: (1) A(Th17) for which only peaks from Th17 48h wildtype conditions were included and (2) A(Th) for which all Th samples were included.

ChIP-seq: TF ChIP-seq and control sequencing data were downloaded from GEO (GSE40918), mapped to the murine genome (mm10) with bowtie2 (2.2.3), filtered based on mapping score (MAPQ > 30, Samtools (0.1.19)), and duplicates removed (Picard). Peaks were called with macs14 version 1.4.2 (parameters: -m 10,30 -g 1865500000 --bw=200) and retained for raw p-value < 10^{-10}. TFs were associated with a target gene, if the ChIP peak fell within +/-10kb of gene body. We tested several peak-gene association rules based on distance from gene body or TSS as well as a raw p-value < 10^{-5} cutoff; downstream methods were robust to these decisions (data not shown).

ENCODE DHS: TF-gene interactions, based on TF-footprinting analysis in ENCODE DNase1-hypersensitivity (DHS) in 25 mouse tissues (Stergachis et al. 2014) were downloaded from http://www.regulatorynetworks.org.

TRRUST: Signed, human TF-gene interactions were downloaded from the Transcriptional Regulatory Relationships Unraveled by Sentence-base Text-mining (TRRUST) database (Han et al. 2015), version 1, and mapped to mouse orthologs using the MGI database (one-to-many mappings included).

ChIP/A(Th17): This prior uses TF ChIP data when available and TF motif analysis of 48h Th17 ATAC-seq data otherwise.

ChIP/A(Th17)+KO+A(Th): This prior represents a combination of ChIP/A(Th17) prior, the 25 TF KO experiments in the G.S., and A(Th) prior. Each prior matrix was Frobenius-norm normalized and multiplied by the number of TFs, so that the predictions of each TF per prior had comparable weight. The interactions in the normalized priors were then summed, and the sign of prior interaction was determined from KO data, when available.

Given the degeneracy of TF motifs, some TFs had identical target gene sets within ATAC-seq and ENCODE DHS priors. Degenerate TFs were merged for TRN inference. For precision-recall analysis, resulting target edges for merged TFs were included for all degenerate TFs in the set.

Inference framework. We used the *Inferelator* algorithm for TRN inference (Bonneau et al. 2006). At steady-state gene expression is modeled as a sparse, multivariate linear combination of TF activities (TFAs):

$$x_{ij} = \sum_{k \in TFs} b_{ik} a_{kj}, \quad \text{[\textbf{Equation 1}]}$$

where x_{ij} corresponds to the expression level of gene i in condition j, a_{jk} is the activity of TF k in condition j, and b_{ik} is describes the effect of TF k on gene i. In the past, a TF's mRNA expression level served as a proxy of protein TF activity. More recently (Arrieta-Ortiz et al. 2015), TFA has been estimated based on partial prior knowledge of a TF's gene targets with the following equation:

$$X = PA, \quad \text{[\textbf{Equation 2}]}$$

where $P \in \mathbb{R}^{|genes| \times |TFs|}$ is the prior matrix of known TF-gene interactions, $X \in \mathbb{R}^{|genes| \times |samples|}$ is the expression matrix for genes in the prior, and $A \in \mathbb{R}^{|TFs| \times |samples|}$ contains the unknown activity levels for TFs in the prior. There is no unique solution to **Equation 2**. The least-squares solution has worked in practice (Arrieta-Ortiz et al. 2015; Tchourine et al. 2017).

For this study, we tested both methods of TFA estimation: (1) TF mRNA levels and (2) based on prior knowledge of target genes [Equation 2]. As described in results, we solved for the interaction terms $\{b_{ik}\}$ in **Equation 1** using the current *Inferelator* engine (BBSR-BIC) as well as a new method (mLASSO-StARS, described in detail below).

Model-building with LASSO and StARS. We used a modified Least Absolute Shrinkage and Selection Operator (mLASSO) to solve for a (sparse) set of TF-gene interaction terms $\{b_{ik}\}$ in **Equation 1**:

$$\hat{B} = argmin_B \, |X - BA|_2^2 + |\Lambda \circ B|_1 \,, \quad \text{[\textbf{Equation 3}]}$$

where X, and A are defined as above, $B \in \mathbb{R}^{|genes| \times |TFs|}$ is the matrix of inferred TF-gene interaction coefficients, $\Lambda \in \mathbb{R}^{|genes| \times |TFs|}$ is a matrix of nonnegative penalties, and $\circ$ represents a Hadamard (entry-wise matrix) product (Studham et al. 2014; Gustafsson et al. 2015). Representing the LASSO penalty as a Hadamard product involving a matrix of penalty terms, as opposed to a single penalty term, enabled us to incorporate prior information into the model-building procedure. Specifically, a smaller penalty Λ_{ik} is used if there is evidence for the TF-gene interaction in the prior matrix. Similar to the G-prior in the current Inferelator BBSR and older Inferelator modified elastic net (MEN) framework (Greenfield et al. 2013a), this procedure encourages selection of interactions supported by the prior (e.g., containing ATAC-seq evidence), if there is also support in the gene expression data. For this study, the entries of the Λ matrices were limited to two values: the nonnegative value λ, for TF-gene interactions without evidence in the prior, and $\text{bias}*\lambda$, where $\text{bias} \in [0,1]$, for TF-gene interactions with support in the prior.

To choose the λ parameter, we tested the Stability Approach to Regularization Selection (StARS) (Liu et al. 2010). The method worked well for a closely related inference problem, functional gene

networks (Zhang et al. 2016; Caballe Mestres et al. 2017), but, to our knowledge, this is the first application of StARS to model selection in the context of TRN inference and with incorporation of prior information as in **Equation 3**. StARS was designed to ensure that the inferred network of interactions includes the true set of network interactions with high probability. In contrast, another popular λ selection method, stability selection, seeks to limit false positive rate (Meinshausen et Bühlmann 2010), which, in a biological setting, might be overly conservative (Liu et al. 2010). Thus, we selected StARS over stability selection. The StARS authors describe their method as an "overselect" method, and this was essentially the behavior we desired to obtain gene models sufficiently large to explain more complex mammalian transcriptional patterns. This property was also key to the TF-TF module analysis describe below, as TF-TF modules were difficult to detect from smaller BBSR-BIC networks.

In brief, StARS rests on the definition of edge instabilities. For a fixed value of λ, instabilities are estimated via subsampling and can be interpreted as twice the Bernoulli variance of a subsampled edge or the fraction of times subsample edge predictions disagree. Authors provide advice about how to select a network with an acceptably (small) level of average edge instability and recommended a cutoff = .05. However, given application of StARS in the new setting of TRN inference and a modified LASSO objective function (matrix of penalties Λ), we decided to use out-of-sample gene expression prediction and precision-recall of G.S. interactions, to aid in selecting an appropriate instability cutoff.

We also tested two ways of calculating average instability: for each gene model individually ("per-gene") and for all gene models at a given λ ("network"). We reasoned that "per-gene" estimates might lead to better models for each gene, but the "network" average instability estimates would be more stable, requiring less search time for the λ corresponding to the desired instability cutoff (**Figure S13**). We visualized the distribution of subsampled TF-gene edge frequencies for per-gene and network average instability estimates at cutoffs = .05, .1 and .2 (**Figure S14**). Typically, the edges selected for the final network are those that are stably nonzero (i.e., present), with an instability less than or equal to the cutoff. For both per-gene and network average instability estimates, we noted that the average network size ranged from ~3.5 TFs/gene to ~9 TFs/gene for the default cutoff .05 to .2. Thus, conventional filtering of edges with StARS would lead to very sparse models, especially using the recommended .05 cutoff. In subsequent analyses, we use StARS to select λ, but not to select a final network cutoff.

We next tested how average instability calculation and cutoff would affect precision-recall and out-of-sample prediction. As we had done previously with application of StARS to ecological network inference (Kurtz et al. 2015), interactions were ranked for precision-recall according to nonzero subsamples per edge (**Figure S14**). However, this lead to a plateau at the beginning of the precision-recall curve (at high precision, low recall) for 50 subsamples (**Figure S15**), and the plateau was not much improved by more time-consuming calculations involving 100 or 200 subsamples. To better and more efficiently prioritize high-confidence edges, we devised the following heuristic:

$$\text{Confidence}(i,k) = \text{Nonzero Subsamples} + |pcorr(i,k)|, \quad [\textbf{Equation 4}]$$

where i and k, again correspond to gene i and TF k, and $pcorr(i,k)$ is the partial correlation between gene i and TF k, for a set model size. (Here, model size was set to 20 TFs/gene for partial correlation estimates; results were not sensitive to a range of sizes: 5-20 TFs/gene.) For TF-gene interactions with the same number of nonzero subsamples, the TF-gene interaction with high absolute partial correlation will be higher confidence. This heuristic had the desired effect for both KO+ChIP (**Figure S15-18**) and KO (**Figure S17**) gold standards (using the Th17 ATAC prior

with prior-based or TF mRNA TFA, no, moderate or strong prior reinforcement, network- and gene-level average instabilities at cutoffs .05, .1 and .2). We also tested the same sets of conditions for all three out-of-sample prediction leave-out sets; representative results for the Early Th17 conditions are shown in **Figure S18**. Overall, precision-recall and out-of-sample prediction were robust to choice of network- versus per-gene average instability and instability cutoff (**Figures S16-18**). Out-of-sample prediction was marginally better using the instability cutoff .05 (**Figure S18**). Because .05 is also the recommended and commonly used cutoff, we selected this parameter. Given that network-level average instability was computationally less expensive than per-gene, we used network-level estimates for all subsequence analyses.

mLASSO-StARS was programmed in MATLAB R2016b, and code relies on the Glmnet for Matlab package (Qian et al. 2013) to solve **Equation 3**. To expedite instability calculations, we use average instability upper and lower bounds from bStARS (Müller et al. 2016). Using two subsamples to calculate upper and lower bounds and 50 subsamples for final instability estimates, the speed up from bStARS was negligible, while lower bounds estimated from five subsamples were tighter (**Figure S19**) and computation time was decreased > two-fold (from 75 to 35 minutes using a single core on a 2017 MacBook Pro with 2.9 GHz Intel Core i7 processor). Code is available at https://github.com/emiraldi/infTRN_lassoStARS.git.

We tested several levels of prior reinforcement for both BBSR-BIC and mLASSO-StARS. For BBSR-BIC, G-prior weights of 1, 1.1, and 1.5 corresponded to no, moderate and strong levels of prior reinforcement, while, for mLASSO-StARS, $\text{bias} = 1,\ .5,\ .25$, (defined above) corresponded to no, moderate and strong reinforcement. These levels of prior reinforcement are similar between BBSR-BIC and mLASS-StARS in terms of resulting prior edges in the TRN (**Figure S2**).

Gene Expression Prediction. As described in **Results**, we generated three leave-out (test) datasets: "Early Th17" (all Th17 time points between 1-16 hours, 8 samples), "All Th0" (Th0 samples for all time points and perturbations, 53 samples), or "Late Th17" (18 Th17 samples from 60-108h post-TCR stimulation) (**Figure 3A**). The samples corresponding to each leave-out set are listed in **Table S4**. For each leave-out prediction challenge, the training set, as typical, was defined as all samples excluding test. For each training set, we performed **both** model selection and parameter estimation independently of the test set. Both BBSR-BIC and the mLASSO-StARS methods provide confidence estimates for predicted TF-gene interactions, and we built TRN models of various sizes as a function of edge confidence cutoffs for each of the training sets. For parameter estimation, training TFA matrices were mean-centered and variance-normalized according to the training-set means $\bar{a}_{train} \in \mathbb{R}^{|TFs|}$ and standard deviations $\bar{\sigma}^{a}_{train} \in \mathbb{R}^{|TFs|}$. Target gene expression vectors were mean-centered according to the training-set mean $\bar{x}_{train} \in \mathbb{R}^{|genes|}$. Then, for each confidence-level cutoff, we regressed the vector of normalized training gene expression data onto the reduced set of normalized training TFA estimates to arrive at a set of multivariate linear coefficients $B_{train} \in \mathbb{R}^{|genes|\times|TFs|}$. Sum of Square Error of prediction was calculated as follows:

$$SSE_{pred} = \sum_{\substack{i\epsilon|genes| \\ j\epsilon\{test\}}} \left(x_{ij} - \sum_{k\epsilon|TFs|} b_{ik,train} \left(\frac{a_{kj} - \bar{a}_{k,train}}{\sigma^{a}_{k,train}} \right) - \bar{x}_{i,train} \right)^2 \quad \text{[Equation 5a]}$$

The "null" model SSE is calculated relative to the mean of training data:

$$SSE_{null} = \sum_{\substack{i\epsilon|genes| \\ j\epsilon\{test\}}} \left(x_{ij} - \bar{x}_{i,train} \right)^2 \quad \text{[Equation 5b]}$$

We also calculated R^2_{pred}, which is a normalized measure of predictive performance, directly related to SSE_{pred} and SSE_{null}:

$$R^2_{pred} = 1 - \frac{SSE_{pred}}{SSE_{null}},\ R^2_{pred} \in (-\infty, 1] \quad \text{[Equation 5c]}$$

Intuitively, the R^2_{pred} compares prediction of the TRN model SSE_{pred} (numerator) to the simplest model (based on mean of observed (training) data, SSE_{null}, denominator). Any R^2_{pred}>0 indicates that the numerator model has some predictive benefit over the null model. For example, R^2_{pred}=.5 (.66) means that the numerator model explains twice (thrice) more data variance than the intercept model.

Final TRNs and De Novo Th17 Core. Given the complementary performance of TF-mRNA- and prior-based TFA, we sought to combine the TRNs using both TFA methods (at moderate prior reinforcement). Given, their unique TF-specific AUPR profile (**Figure 2C**), we combined methods by taking the maximum to preserve the individual strengths of each (Kittler et al. 1996; Castro et al.) and compared to rank-sum (or average) combination (Marbach et al. 2012), which would emphasize commonalities but potentially mask individual strengths. As anticipated, max-rank combination had better precision-recall than rank-sum combination (**Figure S20**). Given the orthogonal nature of the TRRUST prior and that is publically available to those considering ATAC-RNA-seq experimental designs, we also tested rank-combination of ATAC and TRRUST TRNs (combining a total of four TRNs, TF mRNA and prior-based TFA for each prior), but observed no benefit, in terms of improved precision-recall (**Figure S21**). More sophisticated methods for integrating TRNs would likely show benefit over the simple techniques tested here (Parisi et al. 2014; Castro et al. 2018), but are beyond the scope of this work.

De Novo Core Th17 TFs were limited to TFs specifically promoting Th17 gene expression patterns. TFs were included in the core if they met one of two criteria: (1) The TF promotes Th17 gene expression through activation (i.e., the TF's positive edges are enriched in up-regulated Th17 genes at an FDR = 1%) or (2) repression of non-Th17 genes (i.e., TF's negative edges are enriched in down-regulated Th17 genes at an FDR = 1%). The hypergeometric CDF and Benjamini-Hochberg procedure were used to estimate adjusted p-values.

Gold Standards. For the gold standards from our lab, we used recommended cutoffs of .75, .75, and 1.5 for KO, ChIP, and KO+ChIP networks, respectively (Ciofani et al. 2012). For the six additional TF KO experiments, we downloaded networks without filtering (Yosef et al. 2013). For both gold standards, gene symbols were mapped from mm9 to mm10, and only genes mapping to both genome builds were considered in precision-recall analysis.

Network Visualization.

Networks were visualized using a newly designed interactive interface, based on iPython and packages: igraph, numpy and scipy. The interface is still under development and will be detailed in a future publication; software is currently available at https://github.com/simonsfoundation/jp_gene_viz.

TF-TF Module Analysis.

We calculated the number of shared target genes between each pair of TFs, analyzing positive and negative target edges separately. (Edges with |partial correlation| < .01 were excluded from analysis, as were TFs with fewer than 20 gene targets.) TFs vary greatly by number of target genes (**Figures S6, S7**), so we devised an overlap normalization scheme that controlled for the variable number of targets per TF. Specifically, we took inspiration from context likelihood of relatedness (CLR), a background-normalized mutual information score (Faith et al. 2007). We define the background-normalized overlap score z_{ij} between TF i and TF j as:

$$z_{ij} = \begin{cases} \sqrt{{z_{i|j}}^2 + {z_{j|i}}^2}, & if\ z_{i|j}, z_{j|i} > 0 \\ 0, & otherwise \end{cases}, \quad \text{[\textbf{Equation 6}]}$$

where $z_{i|j}$ is the z-score of the overlap between TF i and TF j, using the mean and standard deviation associated with the overlaps of TF j to calculate the z-score. Note that the score is only nonzero if the overlap is above average for both TFs. The normalized overlap score has good agreement with overlap significances estimated using the hypergeometric CDF.

We filtered the normalized overlap matrix so that it contained only TFs with at least one significant overlap (FDR = 10%, hypergeometric CDF). We then converted the similarity matrix of normalized overlaps to a distance matrix for hierarchical clustering using Ward distance. To arrive at a final number of clusters, we calculated the mean silhouette score for solutions over a range of total clusters and selected the solution that maximized mean silhouette score. For positive interactions (149 or 167 TFs), this analysis lead to 42 clusters, and 9 or 11 clusters for negative interactions (31 or 30 TFs), for final ChIP+KO+ATAC or ATAC-only TRNs, respectively.

To prioritize and rank TF-TF clusters, we developed a method based on combination of p-values generated from the empirical distribution of normalized overlap scores. For each cluster, we combined p-values for the pairwise normalized overlap scores between cluster TFs (e.g., for cluster of size s, there were $s(s-1)/2$ pairwise scores). We combined p-values with the weighted z-method (Whitlock 2005). P-values are converted via the inverse normal CDF to z-score space and combined with the following equation:

$$Z_w = \frac{\sum_{i=1}^{n} w_i Z_i}{\sqrt{\sum_{i=1}^{n} w_i}}, \quad \text{[\textbf{Equation 7}]}$$

where n is the number of p-values to be combined and $w \in \mathbb{R}^n$ is a vector of weights. Z_w is then converted to a p-value using the normal CDF. Here we set weights to $2/(s-1)$. The resulting p-value score had the desired properties, ranking clusters on the basis of size and strength of overlap (**Figure S22A**) and was used to prioritize a limited set of 15 positive-edge TF-TF clusters for **Figures 6C, S9, S11**. The significance of negative TF-TF clusters were orders of magnitude smaller than the top-15 positive TF-TF clusters (**Figure S22B**), so were not analyzed further. The lack of significant negative TF-TF clusters might have to do with the observed positive to negative edge bias in our method (~1.8-1.9:1).

Data Access

The newly generated RNA-seq and ATAC-seq data will be made available from NCBI's GEO Database.

Acknowledgements

We thank the Flatiron Institute Scientific Computing Core (I. Fisk) for enabling the computational aspects of this work and the New York University Langone Medical Center Genomics Core (A. Heguy and P. Zappile) for help with sequencing. We are thankful to T. Aijo, M. Weirauch and R. Taylor for advice on the manuscript. This work was supported by the Cincinnati Children's Research Foundation (Trustee Award Grant to E.R.M.), the Simons Foundation (E.R.M., A.W., N.D., N.C., R.B.), U.S. National Institute of Health (P01NS076410 to D.L.; 5T32AI100853 to M.P.; R01-DK103358-01 to R.B. and D.L.; and R01-GM112192-01 to R.B.), Crohn's and Colitis Foundation of America (fellowship to M.C.), Damon Runyon Cancer Research Foundation (Dale and Betty Frey Fellowship to J.A.H.), the Laura and Isaac Perlmutter Cancer Center (P30CA016087 to A.H.).

Disclosure Declaration

The authors have no conflicts of interest.

References

Äijö T, Bonneau R. 2016. Biophysically Motivated Regulatory Network Inference: Progress and Prospects. *Hum Hered* **81**: 62-77.

Arrieta-Ortiz ML, Hafemeister C, Bate AR, Chu T, Greenfield A, Shuster B, Barry SN, Gallitto M, Liu B, Kacmarczyk T, et al. 2015. An experimentally supported model of the Bacillus subtilis global transcriptional regulatory network. *Mol Syst Biol* **11**: 839.

Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan ET, Metzler G, Vedenko A, Chen X. 2009. Diversity and complexity in DNA recognition by transcription factors. *Science (80-)* **324**: 1720-1723.

Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. 2007. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **129**: 823-837.

Beagrie RA, Scialdone A, Schueler M, Kraemer DCA, Chotalia M, Xie SQ, Barbieri M, de Santiago I, Lavitas L-M, Branco MR, et al. 2017. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**: 519.

Blatti C, Kazemian M, Wolfe S, Brodsky M, Sinha S. 2015. Integrating motif, DNA accessibility and gene expression data to build regulatory maps in an organism. *Nucleic Acids Res* **43**: 3998-4012.

Bonneau R, Facciotti MT, Reiss DJ, Schmid AK, Pan M, Kaur A, Thorsson V, Shannon P, Johnson MH, Bare JC, et al. 2007. A Predictive Model for Transcriptional Control of Physiology in a Free Living Cell. *Cell* **131**: 1354-1365.

Bonneau R, Reiss DJ, Shannon P, Facciotti M, Hood L, Baliga NS, Thorsson V. 2006. The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biol* **7**: 1.

Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE. 2008. High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**: 311-322.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**: 1213-8.

Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY, Greenleaf WJ. 2015. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**: 486-90.

Caballe Mestres A, Bochkina N, Mayer C. 2017. Selection of the Regularization Parameter in Graphical Models using Network Characteristics. *J Comput Graph Stat* 0-0.

Castro DM, Veaux N De, Miraldi E, Bonneau R. Joint Inference of Gene Regulatory Networks from Multiple Expression Datasets Improves Model Accuracy.

Castro DM, Veaux N de, Miraldi ER, Bonneau R. 2018. Multi-study inference of regulatory networks for more accurate models of gene regulation. *bioRxiv* 279224.

Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur Ö, Anwar N, Schultz N, Bader GD, Sander C. 2010. Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res* **39**: D685-D690.

Chai LE, Loh SK, Low ST, Mohamad MS, Deris S, Zakaria Z. 2014. A review on the computational approaches for gene regulatory network construction. *Comput Biol Med* **48**: 55-65.

Chen C, Grennan K, Badner J, Zhang D, Gershon E, Jin L, Liu C. 2011. Removing batch effects in analysis of expression microarray data: An evaluation of six batch adjustment methods. *PLoS One* **6**.

Chen X, Yu B, Carriero N, Silva C, Bonneau R. 2017. Mocap: Large-scale inference of transcription factor binding sites from chromatin accessibility. *Nucleic Acids Res* **45**:

4315-4329.
Christie D, Zhu J. 2014. Transcriptional regulatory networks for CD4 T cell differentiation. Dans *Transcriptional Control of Lineage Differentiation in Immune Cells*, p. 125-172, Springer.
Ciofani M, Madar A, Galan C, Sellars M, Mace K, Pauli F, Agarwal A, Huang W, Parkurst CN, Muratet M, et al. 2012. A validated regulatory network for Th17 cell specification. *Cell* **151**: 289-303.
Consortium GO. 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* **32**: D258-D261.
Derrien T, Estellé J, Sola SM, Knowles DG, Raineri E, Guigó R, Ribeca P. 2012. Fast computation and applications of genome mappability. *PLoS One* **7**: e30377.
Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15-21.
Duren Z, Chen X, Jiang R, Wang Y, Wong WH. 2017. Modeling gene regulation from paired expression and chromatin accessibility data. *Proc Natl Acad Sci* **114**: E4914-E4923.
Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS. 2007. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS Biol* **5**: e8.
Fu Y, Jarboe LR, Dickerson JA. 2011. Reconstructing genome-wide regulatory network of E. coli using transcriptome data and predicted transcription factor activities. *BMC Bioinformatics* **12**: 233.
Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res* **17**: 877-885.
Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**: 1017-1018.
Greenfield A, Hafemeister C, Bonneau R. 2013a. Robust data-driven incorporation of prior knowledge into the inference of dynamic regulatory networks. *Bioinformatics* **29**: 1060-1067.
Greenfield A, Hafemeister C, Bonneau R. 2013b. Robust data-driven incorporation of prior knowledge into the inference of dynamic regulatory networks. *Bioinformatics* **29**: 1060-1067.
Gustafsson M, Gawel DR, Alfredsson L, Baranzini S, Bjorkander J, Blomgran R, Hellberg S, Eklund D, Ernerudh J, Kockum I, et al. 2015. A validated gene regulatory network and GWAS identifies early regulators of T cell-associated diseases. *Sci Transl Med* **7**: 1-10.
Han H, Shim H, Shin D, Shim JE, Ko Y, Shin J, Kim H, Cho A, Kim E, Lee T, et al. 2015. TRRUST: a reference database of human transcriptional regulatory interactions. *Sci Rep* **5**: 11432.
Hecker M, Lambeck S, Toepfer S, van Someren E, Guthke R. 2009. Gene regulatory network inference: Data integration in dynamic models-A review. *BioSystems* **96**: 86-103.
Heng TSP, Painter MW, Elpek K, Lukacs-Kornek V, Mauermann N, Turley SJ, Koller D, Kim FS, Wagers AJ, Asinovski N. 2008. The Immunological Genome Project: networks of gene expression in immune cells. *Nat Immunol* **9**: 1091.
Isono F, Fujita-Sato S, Ito S. 2014. Inhibiting RORγt/Th17 axis for autoimmune disorders. *Drug Discov Today* **19**: 1205-1211.
Jolma A, Kivioja T, Toivonen J, Cheng L, Wei G, Enge M, Taipale M, Vaquerizas JM, Yan J, Sillanpa MJ, et al. 2010. Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. 861-873.
Jolma A, Yan J, Whitington T, Toivonen J, Nitta KR, Rastas P, Morgunova E, Enge M, Taipale M, Wei G, et al. 2013. DNA-binding specificities of human transcription factors. *Cell* **152**: 327-339.

Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**: 27-30.

Karwacz K, Miraldi ER, Pokrovskii M, Madi A, Yosef N, Wortman I, Chen X, Watters A, Carriero N, Awasthi A, et al. 2017. Critical role of IRF1 and BATF in forming chromatin landscape during type 1 regulatory cell differentiation. *Nature* **18**: 412-421.

Khader SA, Gaffen SL, Kolls JK. 2009. Th17 cells at the crossroads of innate and adaptive immunity against infectious diseases at the mucosa. *Mucosal Immunol* **2**: 403-411.

Kheradpour P, Kellis M. 2014. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res* **42**: 2976-2987.

Kittler J, Hater M, Duin RPW. 1996. Combining classifiers. *Proc - Int Conf Pattern Recognit* **2**: 897-901.

Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. 2015. Sparse and Compositionally Robust Inference of Microbial Ecological Networks. *PLoS Comput Biol* **11**: e1004226.

Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR, Weirauch MT. 2018. The Human Transcription Factors. *Cell* **172**: 650-665.

Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, et al. 2002. Transcriptional Regulatory Networks in Saccharomyces cerevisiae *Science (80-)* **298**: 799 LP-804.

Li P, Spolski R, Liao W, Leonard WJ. 2014. Complex interactions of transcription factors in mediating cytokine biology in T cells. *Immunol Rev* **261**: 141-156.

Liao JC, Boscolo R, Yang Y-L, Tran LM, Sabatti C, Roychowdhury VP. 2003. Network component analysis: reconstruction of regulatory signals in biological systems. *Proc Natl Acad Sci U S A* **100**: 15522-7.

Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. 2011. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**: 1739-1740.

Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (80-)* **326**: 289-293.

Littman DR, Rudensky AY. 2010. Th17 and Regulatory T Cells in Mediating and Restraining Inflammation. *Cell* **140**: 845-858.

Liu H, Roeder K, Wasserman L. 2010. Stability Approach to Regularization Selection (StARS) for High Dimensional Graphical Models. Dans *Advances in Neural Information Processing Systems 23* (éd. J.D. Lafferty, C.K.I. Williams, J. Shawe-Taylor, R.S. Zemel, et A. Culotta), p. 1432-1440, Curran Associates, Inc.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

Madar A, Greenfield A, Vanden-Eijnden E, Bonneau R. 2010. DREAM3: Network inference using dynamic context likelihood of relatedness and the inferelator. *PLoS One* **5**.

Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, Allison KR, Aderhold A, Bonneau R, Chen Y. 2012. Wisdom of crowds for robust gene network inference. *Nat Methods* **9**: 796.

McCarthy MT, O'Callaghan CA. 2014. PeaKDEck: a kernel density estimator-based peak calling program for DNaseI-seq data. *Bioinformatics* **30**: 1302-1304.

Meinshausen N, Bühlmann P. 2010. Stability selection. *J R Stat Soc Ser B (Statistical Methodol* **72**: 417-473.

Müller CL, Bonneau R, Kurtz Z. 2016. Generalized Stability Approach for Regularized Graphical Models.

Najafabadi HS, Mnaimneh S, Schmitges FW, Garton M, Lam KN, Yang A, Albu M, Weirauch MT, Radovani E, Kim PM. 2015. C2H2 zinc finger proteins greatly expand the human

regulatory lexicon. *Nat Biotechnol* **33**: 555.
Neph S, Stergachis AB, Reynolds A, Sandstrom R, Borenstein E, Stamatoyannopoulos JA. 2012. Circuitry and Dynamics of Human Transcription Factor Regulatory Networks. *Cell* **150**: 1274-1286.
Ouyang Z, Zhou Q, Wong WH. 2009. ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells. *Proc Natl Acad Sci* **106**: 21521-21526.
Parisi F, Strino F, Nadler B, Kluger Y. 2014. Ranking and combining multiple predictors without labeled data. *Proc Natl Acad Sci U S A* **111**: 1253-8.
Patel DD, Kuchroo VK. 2015. Th17 Cell Pathway in Human Immunity: Lessons from Genetics and Therapeutic Interventions. *Immunity* **43**: 1040-1051.
Pico AR, Kelder T, Van Iersel MP, Hanspers K, Conklin BR, Evelo C. 2008. WikiPathways: pathway editing for the people. *PLoS Biol* **6**: e184.
Pique-Regi R, Degner JF, Pai AA, Gaffney DJ, Gilad Y, Pritchard JK. 2011. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res* **21**: 447-455.
Qian J, Hastie T, Friedman J, Tibshirani R, Simon N. 2013. Glmnet for Matlab.
Qin J, Hu Y, Xu F, Yalamanchili HK, Wang J. 2014. Inferring gene regulatory networks by integrating ChIP-seq/chip and transcriptome data via LASSO-type regularization methods. *Methods* **67**: 294-303.
Ramirez RN, El-Ali NC, Mager MA, Wyman D, Conesa A, Mortazavi A. 2017. Dynamic Gene Regulatory Networks of Human Myeloid Differentiation. *Cell Syst* **4**: 416-429.e3.
Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, et al. 2000. Genome-Wide Location and Function of DNA Binding Proteins. *Science (80-)* **290**: 2306 LP-2309.
Rendeiro AF, Schmidl C, Strefford JC, Walewska R, Davis Z, Farlik M, Oscier D, Bock C. 2016. Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks. *Nat Commun* **7**.
Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, et al. 2007. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* **4**: 651.
Sherwood RI, Hashimoto T, O'Donnell CW, Lewis S, Barkal AA, Van Hoff JP, Karun V, Jaakkola T, Gifford DK. 2014. Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat Biotechnol* **32**: 171-178.
Siahpirani AF, Roy S. 2016. A prior-based integrative framework for functional transcriptional regulatory network inference. *Nucleic Acids Res* **45**: gkw963.
Stadhouders R, Lubberts E, Hendriks RW. 2017. A cellular and molecular view of T helper 17 cell plasticity in autoimmunity. *J Autoimmun*.
Stergachis AB, Neph S, Sandstrom R, Haugen E, Reynolds AP, Zhang M, Byron R, Canfield T, Stelhing-Sun S, Lee K, et al. 2014. Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature* **515**: 365-370.
Studham ME, Tjärnberg A, Nordling TEM, Nelander S, Sonnhammer ELL. 2014. Functional association networks as priors for gene regulatory network inference. *Bioinformatics* **30**: 130-138.
Tchourine K, Vogel C, Bonneau R. 2017. Explicit Modeling of RNA Stability Improves Large-Scale Inference of Transcription Regulation. *bioRxiv*.
Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, et al. 2014. Determination and Inference of Eukaryotic Transcription Factor Sequence Specificity. *Cell* **158**: 1431-1443.
Whitlock MC. 2005. Combining probability from independent tests: the weighted Z-method is

superior to Fisher's approach. *J Evol Biol* **18**: 1368-73.
Wilkins O, Hafemeister C, Plessis A, Holloway-Phillips M-M, Pham GM, Nicotra AB, Gregorio GB, Jagadish SVK, Septiningsih EM, Bonneau R, et al. 2016. EGRINs (Environmental Gene Regulatory Influence Networks) in Rice That Function in the Response to Water Deficit, High Temperature, and Agricultural Environments. *Plant Cell* **28**: 2365-2384.
Wingender E. 2008. The TRANSFAC project as an example of framework technology that supports the analysis of genomic regulation. *Brief Bioinform* **9**: 326-332.
Wingender E, Schoeps T, Haubrock M, Dönitz J. 2014. TFClass: a classification of human transcription factors and their rodent orthologs. *Nucleic Acids Res* **43**: D97-D102.
Xi H, Shulha HP, Lin JM, Vales TR, Fu Y, Bodine DM, McKay RDG, Chenoweth JG, Tesar PJ, Furey TS. 2007. Identification and characterization of cell type–specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS Genet* **3**: e136.
Yang J, Sundrud MS, Skepner J, Yamagata T. 2014. Targeting Th17 cells in autoimmune diseases. *Trends Pharmacol Sci* **35**: 493-500.
Yosef N, Shalek AK, Gaublomme JT, Jin H, Lee Y, Awasthi A, Wu C, Karwacz K, Xiao S, Jorgolli M, et al. 2013. Dynamic regulatory network controlling TH17 cell differentiation. *Nature* **496**: 461-8.
Zhang HM, Liu T, Liu CJ, Song S, Zhang X, Liu W, Jia H, Xue Y, Guo AY. 2015. AnimalTFDB 2.0: A resource for expression, prediction and functional study of animal transcription factors. *Nucleic Acids Res* **43**: D76-D81.
Zhang J, Poh HM, Peh SQ, Sia YY, Li G, Mulawadi FH, Goh Y, Fullwood MJ, Sung W-K, Ruan X. 2012. ChIA-PET analysis of transcriptional chromatin interactions. *Methods* **58**: 289-299.
Zhang XF, Ou-Yang L, Zhao XM, Yan H. 2016. Differential network analysis from cross-platform gene expression data. *Sci Rep* **6**: 1-13.

Supplemental Figure Legends

Figure S1. PCA of Gene Expression Profiles. Upper panels show scores plots, where the 254 RNA-seq samples are plotted as a function of all genes for principal components (PCs) 1-4. Lower panels display gene loadings for the top-75 gene contributors to PCs 1-4.

Figure S2. The effect of TFA and prior reinforcement to recovery of prior edges in TRN models. % prior edges in TRN are plotted as a function of TRN size (mean TF predictors per gene) for the two Inferelator methods (BBSR-BIC and mLASSO-StARS), the two methods for estimating TF activities (TFA = P^+X and TF mRNA), and varying levels of prior reinforcement (-, +, ++ correspond to no, moderate, and high prior reinforcement, respectively).

Figure S3. Precision-recall of multiple priors, Gold Standard = KO+ChIP (A) or KO (B). For each prior, the performance is plotted for several TRNs, based on Inferelator method (LS = mLASSO-StARS (reds), BB = BBSR-BIC (blues)), TFA estimate (m = TF mRNA, TFA = P^+X), and "+" corresponds to strength of prior reinforcement. Random and "No Prior" control TRNs serve as references in all panels.

Figure S4. Precision-recall of multiple priors, G.S. = KO+ChIP (A) or KO (B), effect of RNA-seq sample size. Left and right panels show performance of TRNs built from 254 and 50 RNA-seq samples, respectively. For each prior, the performance is plotted for several TRNs, based on Inferelator method (LS = mLASSO-StARS (reds), BB = BBSR-BIC (blues)), TFA estimate (m = TF mRNA, TFA = P^+X), and "+" corresponds to strength of prior reinforcement. Random and "No Prior" control TRNs serve as references in all panels.

Figure S5. Out-of-sample prediction of gene expression patterns extended to include TFA = TF mRNA and No Prior TRNs. Leave-out sets are defined as in **Figure 4A**. Model selection and parameter estimation of TF-gene interactions were performed in the absence of specified leave-out sets (All Th0, Late Th17, and Early Th17). **(A) and (B) Gene expression prediction.** TRN models were built without prior (left panels), with an ATAC (central panels) or ChIP+KO+ATAC priors (right panels), using mLASSO-StARS, bias = .5, and TFA = TF mRNA (upper panels) TFA = P^+X (lower panels). R^2_{pred} for each leave-out set is plotted as a function of model instability cutoff **(A)** or average number of TFs per gene **(B)**. In the key, the size of each leave-out set is denoted in parentheses. For reference, precision, recall, and F1-score of the full model are also plotted, providing intuition about the number of KO or KO+ChIP edges are in the TRN at each cut off.

Figure S6. Additional Th17 core TRN models. "Core" Th17 genes and TFs were selected from the literature for visual comparison with jp_gene_viz software. Network size was limited to an average of 15 TFs per gene. Source of network and TFA method is denoted for each network. "+" denote moderate prior reinforcement with bias = .5. The edges in Inferelator TRNs are colored according to partial correlation (red positive, blue negative) and weighted relative to edge stability. For the full KO-ChIP G.S. from Ciofani et al., edge sign is based on differential gene expression analysis between TF KO and control and edge weight is constant. Nodes are colored according to gene expression at 48h in Th17, z-scored relative to Th17 and other T Helper cell time points (red/blue = increased/decreased expression).

Figure S7. Distribution of targets per TF in final ChIP+KO+ATAC TRN. Network size was limited to mean 15 TFs per gene and network edges were further filter to remove any edge with absolute partial correlation < .01, leading to ~80% reduction in network size. TFs were ranked according to degree (total number of target genes) in the left panel, while the inset displays the

Top 100 highest degree TFs. TF target genes are colored according to interaction sign and whether the interaction was also in the prior (see Key). Asterisks denote TFs within the Top-20 highest degree in both ChIP+KO+ATAC and ATAC TRNs.

Figure S8. Distribution of targets per TF in final ATAC TRN. Network size was limited to mean 15 TFs per gene and network edges were further filter to remove any edge with absolute partial correlation < .01, leading to ~80% reduction in network size. TFs were ranked according to degree (total number of target genes) in the left panel, while the inset displays the Top 100 highest degree TFs. TF target genes are colored according to interaction sign and whether the interaction was also in the prior (see Key). Asterisks denote TFs within the Top-20 highest degree in both ChIP+KO+ATAC and ATAC TRNs.

Figure S9. Top 15 TF modules for ChIP+KO+ATAC TRN. TFs were clustered into modules on the basis of shared positive target genes between TFs (see **Methods**). The left panel shows enrichment of positive TF target interactions with up-regulated Th17 genes ("Th17" for potential Th17-promoting TFs) or down-regulated Th17 genes ("Non-Th17", for potential Th17-repressing TFs). The lower panel shows relative TF gene expression over the Th17 and other T Helper cell time points.

Figure S10. Gene-set enrichment of ChIP+KO+ATAC TRN TF modules. We compiled gene sets from five databases: GO, Pathway Commons, KEGG, MAPP and Signatures from MSigDB. For each TF, we calculated the significance of overlap between the TF's target genes and gene sets (hypergeometric CDF). For Kegg and Mapp databases, the color bar maximum is 7, and, for the other database, the maximum is 11. TFs are clustered and color-coded according to the TF-TF modules in **Figure S9, S11**. TF-TF module enrichments are consistent among gene sets and TRNs, providing robust predictions for modules in Th17 biology.

Figure S11. Top 15 TF modules for ATAC-only TRN. TFs were clustered into modules on the basis of shared positive target genes between TFs (see **Methods**). The left panel shows enrichment of positive TF target interactions with up-regulated Th17 genes ("Th17" for potential Th17-promoting TFs) or down-regulated Th17 genes ("Non-Th17", for potential Th17-repressing TFs). The lower panel shows relative TF gene expression over the Th17 and other T Helper cell time points.

Figure S12. Gene-set enrichment of ATAC-only TRN TF modules. We compiled gene sets from five databases: GO, Pathway Commons, KEGG, MAPP and Signatures from MSigDB. For each TF, we calculated the significance of overlap between the TF's target genes and gene sets (hypergeometric CDF). For Kegg and Mapp databases, the color bar maximum is 7, and, for the other database, the maximum is 11. TFs are clustered and color-coded according to the TF-TF modules in **Figure S9, S11**. TF-TF module enrichments are consistent among gene sets and TRNs, providing robust predictions for modules in Th17 biology.

Figure S13. **(A) Per-gene and (B) network-level average instability paths.** Average instabilities were calculated on per-gene and network level over a range of LASSO penalties, λ. Results are shown for the No Prior TRN using 50 subsamples.

Figure S14. **Distribution of nonzero subsamples per TF-gene edge.** Distribution of nonzero subsamples per TF-gene, using per-gene and network average instability cutoffs of .05, .1, and .2. Results are shown for the No Prior TRN using 50 subsamples. The dotted-red line marks the instability cutoff in terms of nonzero subsamples, and the average model size at that cutoff appears as text in the upper right hand.

Figure S15. **Comparison of edge-ranking methods**. Edges are ranked according to **(A)** nonzero subsamples or **(B)** nonzero subsamples plus the absolute value of edge partial correlation (**Equation 4**) for network- and gene-level average instabilities at cutoffs .05, .1 and .2. Right and left panels denote full and zoomed-in curves, respectively. Results are shown for the Th17 ATAC prior, bias = .5, and prior-based TFA, using the KO + ChIP G.S.

Figure S16. **StARS parameterization: precision-recall of the KO+ChIP G.S.** TRNs were built using the Th17 ATAC prior with prior-based or TF mRNA TFA, no, moderate or strong prior reinforcement, network- and gene-level average instabilities at cutoffs .05, .1 and .2. Edges are ranked according to nonzero subsamples or (**Equation 4**).

Figure S17. **StARS parameterization: precision-recall of the KO G.S.** TRNs were built using the Th17 ATAC prior with prior-based or TF mRNA TFA, no, moderate or strong prior reinforcement, network- and gene-level average instabilities at cutoffs .05, .1 and .2. Edges are ranked according to nonzero subsamples or (**Equation 4**).

Figure S18. **StARS parameterization: out-of-sample gene expression prediction.** TRNs were built using the Th17 ATAC prior with prior-based or TF mRNA TFA, no, moderate or strong prior reinforcement, network- and gene-level average instabilities at cutoffs .05, .1 and .2. Out-of-sample gene expression prediction was tested on the leave-out "Early Th17" set (8 samples).

Figure S19. **Computational speed-up with bStARS lambda bounds. (A)** Network instabilities were estimated using 50 subsamples for Th17 ATAC prior with prior-based TFA and moderate reinforcement over a range of lambda penalties containing the lambda corresponding to the target instability cutoff .05 (vertical black line). bStARS upper and lower instability bounds, based on **(B)** two or **(C)** five subsamples shorten the lambda search space (orange arrows, and black lines indicate lower and upper bounds for lambda corresponding to the target instability cutoff .05). Solving the LASSO is very slow for smaller lambda penalties; thus, increasing the lower bound (as in **(C)**) results in significant speed-up (~2-fold), while looser bounds (as in **(B)**) do not.

Figure S20. **Rank-combination of ATAC TRNs.** The precision-recall of individual ATAC TRNs (at moderate prior reinforcement "b=50" for prior-based and TF mRNA TFA) are compared to performance average- or maximum-combination of TRNs.

Figure S21. **Rank-combination of TRNs from different priors.** The precision-recall of TRNs from individual priors and pairs of priors (at moderate prior reinforcement "b=50" for both prior-based and TF mRNA TFA) are compared, using **(A)** maximum- or **(B)** average-combination of TRNs.

Figure S22. **Ranking of TF-TF Clusters.** Heuristic significance estimates versus cluster size for TF-TF modules built from **(A)** positive and **(B)** negative TF-TF edge overlaps.

Supplementary Table Legends

Table S1. Statistics on Priors and Gold Standards. (A) and **(B)** were generated using the initial and updated mouse TF lists, respectively (see **Methods**).

Table S2. RNA-seq Sample Metadata. % prior edges in TRN are plotted as a function of TRN size (mean TF predictors per gene) for the two Inferelator methods (BBSR-BIC and mLASSO-StARS), the two methods for estimating TF activities (TFA = $P^{+}X$ and TF mRNA), and various levels of prior reinforcement.

Table S3. Target Gene and Candidate Regulator Lists. % prior edges in TRN are plotted as a function of TRN size (mean TF predictors per gene) for the two Inferelator methods (BBSR-BIC and mLASSO-StARS), the two methods for estimating TF activities (TFA = $P^{+}X$ and TF mRNA), and various levels of prior reinforcement.

Table S4. Lists of sample names corresponding to each leave-out set.