Causal Network Inference from Gene Transcriptional Time Series Response to Glucocorticoids

Jonathan Lu,[1,*] Bianca Dumitrascu,[2,*] Ian C. McDowell,[3] Brian Jo,[2]
Alejandro Barrera,[4,5], Linda K. Hong,[4] Sarah M. Leichter,[4]
Timothy E. Reddy,[6,†] Barbara E. Engelhardt[1,7,†]

[1] *Department of Computer Science, Princeton University, Princeton, NJ 08540, USA*
[2] *Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA*
[3] *Element Genomics, A UCB Company, Durham, NC, 27701, USA*
[4] *Center for Genomic and Computational Biology, Duke University, Durham, NC 27708, USA*
[5] *Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC 27710, USA*
[6] *Department of Genome Sciences, Duke University, Durham, NC 27708, USA*
[7] *Center for Statistics and Machine Learning, Princeton University, Princeton, NJ 08540, USA*

[*] *These authors contributed equally*
[,†] *To whom correspondence should be addressed; E-mail: bee@princeton.edu, tim.reddy@duke.edu*

Abstract

Gene regulatory network inference is essential to uncover complex relationships among gene pathways and inform downstream experiments, ultimately paving the way for regulatory network re-engineering. Network inference from transcriptional time series data requires accurate, interpretable, and efficient determination of causal relationships among thousands of genes. Here, we develop Bootstrap Elastic net regression from Time Series (BETS), a statistical framework based on Granger causality for the recovery of a directed gene network from transcriptional time series data. BETS uses elastic net regression and stability selection from bootstrapped samples to infer causal relationships among genes. BETS is highly parallelized, enabling efficient analysis of large transcriptional data sets. We show competitive accuracy on a community benchmark, the DREAM4 100-gene network inference challenge, where BETS is one of the fastest among methods of similar performance but additionally infers whether the causal effects are activating or inhibitory. We apply BETS to transcriptional time series data of $2,768$ differentially-expressed genes from A549 cells exposed to glucocorticoids over a period of 12 hours. We identify a network of $2,768$ genes and $31,945$ directed edges (FDR ≤ 0.2). We validate inferred causal network edges using two external data sources: overexpression experiments on the same glucocorticoid system, and genetic variants associated with inferred edges in primary lung tissue in the Genotype-Tissue Expression (GTEx) v6 project. BETS is freely available as an open source software package at `https://github.com/lujonathanh/BETS`.

Keywords: gene regulation, network inference, directed networks, bulk RNA-seq, glucocorticoids, vector autoregression, Granger causality, time series, stability selection

1. Introduction

The recent availability of gene expression measurements over time has enabled the search for interpretable statistical models of gene regulatory dynamics [1]. These time series data present a unique opportunity to use the coordinated transcriptional response to environmental exposure to infer causal relationships between genes. However, there are several challenges to overcome in the analysis of time series transcriptomic data. These data are generally high-dimensional: the number of quantified gene transcripts—approximately 20,000 in human samples—often dramatically exceeds the number of available time points and samples. Many classical statistical assumptions fail to hold in this high-dimensional regime [2, 3]. Moreover, the large number of gene transcripts poses a computational burden, as the number of possible edges in a gene network grows quadratically. Finally, a transcriptional time series often has a small number of time points, and those time points are often not uniformly spaced; furthermore, because transcriptional time series data often quantify transcription post exposure, the time series is not stationary, and genes respond to the exposure and return to baseline at different rates [4, 5].

In this work, we develop an approach that uses the gene transcriptional time series following glucocorticoid (GC) exposure to build a directed gene network [6]. GCs play an essential role in regulating stress response, and are widely used as anti-inflammatory and immunosuppresive medication [6, 7]. Despite clinical benefits, prolonged exposure to GCs has been linked to increased risk for type 2 diabetes mellitus (T2DM) [8] and obesity [9]. Here, we develop a method to accurately, interpretably, and efficiently infer a directed gene network using transcriptional time series data. We focus our analysis of this network on immune-related genes, metabolism-related genes, and transcription factors (TFs) to study the inferred coordinated response of these systems to GCs.

Our method, Bootstrap Elastic net inference from Time Series (BETS), uses vector autoregression with elastic net regularization to accurately infer directed edges between genes. Stability selection, which assesses the robustness of an edge to perturbations in the data, leads to improvements over baseline vector autoregression methods in this high-dimensional context [3]. Furthermore, BETS is biologically interpretable because estimated coefficients provide the direction (sign) and effect size of the causal relationship between a pair of genes. Finally, BETS's parallelization enables efficient inference of networks with millions of possible edges in a computationally tractable way.

We use the causal network inferred by BETS on the GC time series data to study the relationships between TFs, immune genes, and metabolic genes. We validate our network using two approaches: ten measurements of the same GC system with a specific TF overexpressed, and an expression quantitative trait loci (eQTL) study [10]. Although our framework is motivated by transcriptional response to GC exposure, our approaches are general, and BETS is applicable to inferring directed networks from arbitrary transcriptional time series.

2. Related Work

Several methods have been developed to estimate directed gene networks from transcriptional time series data (Figure S1) [11, 12, 13, 14, 15, 16, 17, 18, 19]. These methods estimate directed networks in which the directed edges between nodes—representing genes—indicate a cause-effect relationship between genes, i.e., perturbing expression of the *causal gene* would lead to changes in expression of the *effect gene* [20]. We briefly overview these methods; for detailed discussion, see Supplemental Information. Here, we take g' to be the causal gene and g to be the effect gene, and quantify support for a causal edge $g' \to g$ in the data.

Mutual information (MI) methods assess the MI between the expression of g' at the previous time point and the expression of g at the current time point (Figure S1A) [21, 22, 23, 24, 25, 26]. A causal edge $g' \to g$ is included in the network if the MI of the two genes across time exceeds a threshold.

Granger causality methods determine if including the expression of g' at the previous time point improves our ability to predict the expression of g at the current time point above using the expression of g at the previous time point [27]. A common way to implement Granger causality is through a vector autoregression (VAR) model, which assumes a linear relationship between all genes' expression at the previous time point and the expression of g at the current time point. A causal edge $g' \rightarrow g$ is included in the network when g' has a statistically significant coefficient in the VAR.

Ordinary differential equations (ODEs) fit the derivative of the expression of g as a function of all genes' expression at a single time point (Figure S1C) [11, 28, 29]. ODE methods typically assume linearity, as small sample sizes make it challenging to infer the parameters of nonlinear functions. A causal edge $g' \rightarrow g$ is included when g' has a statistically significant coefficient in the ODE.

Decision trees (DTs) are a type of nonparametric function based on partitioning the data [30, 31]. DT methods fall either under VAR or ODE; either the DTs fit the expression of g at the current time as a function of all genes' expression at the previous time point (VAR), or they fit the derivative of the expression of g as a function of all genes' expression at a single time point (ODE) (Figure S1D) [32, 33]. A causal edge $g' \rightarrow g$ is included in the network when an importance score for g' exceeds some threshold, where importance scores are typically the reduction in variance of g when g' is included as a predictor.

Dynamic Bayesian networks (DBNs) search the space of possible directed acyclic graphs between previous and current expression levels to identify the network structure with the highest posterior probability of each edge given the data (Figure S1E) [34, 35, 36, 37, 38]. DBNs typically assume a linear relationship between previous and current expression. A causal edge $g' \rightarrow g$ is included in the network when its marginal posterior probability of existence exceeds some threshold.

A Gaussian process (GP) is a distribution over continuous, nonlinear functions. GPs are often used in the context of nonlinear DBNs, where GP regression is used to model a nonlinear relationship between previous expression and current expression (Figure S1F) [39, 40]. A causal edge $g' \rightarrow g$ is included in the network based on its posterior probability of existence exceeding some threshold.

3. Results

First, we briefly describe the approach in BETS to infer a directed gene network. Next, we compare results from BETS to those from twenty other methods on the 100-gene time series data from the DREAM4 Network Inference Challenge [41]. Then, we describe the network estimated from the GC transcriptional time series data. Finally, we validate the inferred network using two different frameworks: overexpression experiments on the same system, and genetic variants associated with inferred edges in primary lung tissue in the Genotype-Tissue Expression (GTEx) v6 project [10].

3.1. BETS: A vector autoregressive approach to causal inference of gene regulatory networks.

Directed networks represent causal relationships among diverse interacting variables in complex systems. We developed a robust, scalable approach based on ideas from Granger causality to construct these directed networks from short, high-dimensional time series observations of gene expression levels.

Let G be the set of all $p = |G|$ genes in the data set and $g \in G$ be a gene. Let $\neg g$ be G with g removed. Let t be a single time point, ranging from $\{1, 2, \ldots, T\}$. Let X_t^g be the expression of gene g at time t. Let L be the time lag, or the number of previous time point observations; so $L = 2$ means that we use two previous time points, $t - 1$ and $t - 2$, to predict expression at time t.

Definition 3.1 (Granger causality)**.** For lag L, a gene g' is said to *Granger-cause* another gene g if using $X^{g'}_{t-1}, \ldots, X^{g'}_{t-L}$, the expression value of g' at times $t-1$ to $t-L$, improves prediction of X^{g}_{t}, the expression value of g at time t, beyond the prediction using $X^{g}_{t-1}, \ldots, X^{g}_{t-L}$ alone.

To test for Granger causality from g' to g, we first preprocessed the gene expression time series data (STAR Methods). For every potential effect gene g, we fit all other genes $g' \in \neg g$ simultaneously (Equation 1), echoing ideas from the graphical lasso for undirected network inference [42]. Intuitively, this adapts the idea of Granger causality to conditional Granger causality, where we consider how gene g' Granger causes g conditioning on the effects of all other genes. This approach uses the regression:

$$X^{g}_{t} = \sum_{\ell=1}^{L} \alpha^{g}_{\ell} X^{g}_{t-\ell} + \sum_{g' \in \neg g} \sum_{\ell=1}^{L} \beta^{g',g}_{\ell} X^{g'}_{t-\ell} + \epsilon_t, \tag{1}$$

where $\epsilon_t \sim \mathcal{N}(0,1)$. For BETS, we set $L = 2$. To test for an edge, if $\beta^{g',g}_{\ell} \neq 0$, then we say g' conditionally Granger-causes g at lag ℓ. We build the directed network by including a directed edge to g from every gene g' that has been inferred to conditionally Granger-cause g.

Robustly building this network is difficult due to the high dimensionality of the problem: the number of genes that could Granger-cause a given g far exceeds the available time points and technical replicates. To address this challenge, BETS regularizes the VAR model parameters using an *elastic net* penalty (STAR Methods, Figure 1A). Elastic net regression encourages sparsity and performs automatic variable selection on the genes being tested for causal influence [43]. The elastic net penalty, unlike the lasso penalty [44], is able to select groups of correlated variables and allows the number of selected variables to be greater than the number of samples. This is particularly important for gene expression assays where gene expression levels are often well-correlated and there are far more genes than samples.

In BETS, we fit the same VAR model to a data set in which causal genes have their expression permuted over time to generate a null distribution of edge coefficients. The coefficients are thresholded to produce a causal network with each edge at edge false discovery rate (FDR) ≤ 0.05 (Figure 1A). We then apply this network inference procedure to multiple (here, $1,000$) bootstrapped samples of the original data set (Figure 1B). Each edge has a *selection frequency*, or the frequency that the edge appears in networks inferred from the bootstrapped samples. Inspired by stability selection, this approach assesses if network edges are robust to perturbations of the data [3]. Finally, we run this overall procedure on a permuted version of the original data set to obtain a null distribution of selection frequencies (Figure 1C). The selection frequency threshold for including each edge is chosen to control the stability FDR ≤ 0.2. As a baseline, we compare BETS against Enet, which runs elastic net regression without stability selection to produce a causal network with each edge at edge FDR ≤ 0.05 (Figure 1A).

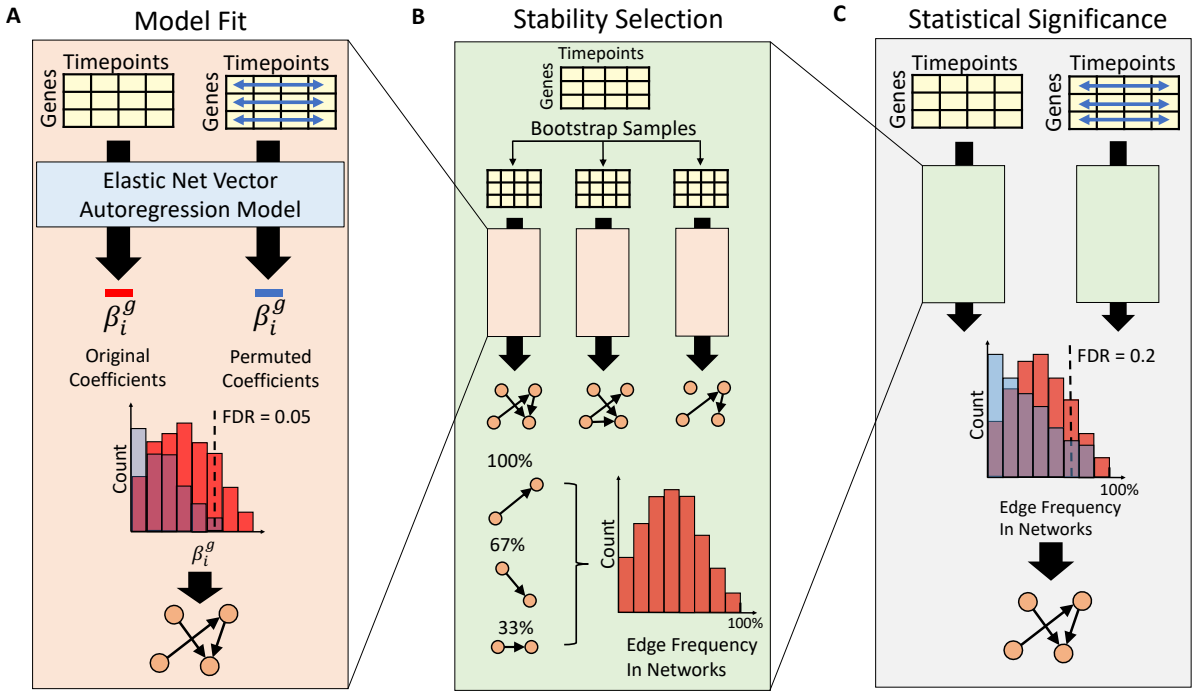

Figure 1: **BETS Algorithm. A) Model Fit.** The VAR model is fit on both the original and a permuted data set (blue arrows indicate shuffling each gene's expression independently across time). Based on the null distribution of coefficients, a threshold is chosen to control the edge FDR at ≤ 0.05. **B) Stability Selection.** From the original data, 1,000 bootstrap samples are generated. For each sample, a network is inferred as in A. Each edge's selection frequency across the bootstrapped networks is computed. **C) Statistical Significance.** For both the original and permuted data, a selection frequency distribution is generated for stability selection as in B. Edges are thresholded to control the stability FDR at ≤ 0.2. See also Figure S1 for an overview of network inference methods.

3.2. Leading Performance on DREAM Network Inference Challenge.

We evaluated BETS against other directed network inference methods. We used the DREAM4 Network Inference Challenge [41], a community benchmark for directed network inference using gene time series data. This benchmark consisted of five data sets, each with ten time series measurements for 100 genes across 21 time points [41]. Evaluation was previously done by looking at the average of the area under the precision recall curve (AUPR) or the area under the receiver operating characteristic (AUROC) over the five data sets [33, 41]. Any method that provides a ranking of possible network edges could be evaluated in this framework.

We tested BETS and Enet against 20 other methods on the DREAM challenge [32, 33, 36, 45, 46]. We ran CSId, Jump3, CLR, MRNET, and ARACNE in-house and found our results consistent with those reported in the literature. All 20 methods reported AUPR, but only 15 reported AUROC.

BETS ranked 6th out of 22 in AUPR with an average AUPR of 0.128 (Figure 2A, Table S1) and 3rd out of 17 in AUROC with an average AUROC of 0.688 (Figure 2B, Table S2). BETS was the top performer of all VAR methods, and Enet was second best. All 22 methods outperformed random selection of edges, which achieved an average AUPR of 0.002 and average AUROC of 0.50 [45]. We also found that BETS and Enet had similar performance to the DBN methods in AUPR, and outperformed most of them in AUROC. Ranked by the top AUPR of each class of methods, the best performing class was GP, followed by DT, MI, VAR,

DBN, and ODE [32, 36, 45]. The VAR method used in BETS produces edge signs (indicating excitatory or inhibitory causal effects) and effect sizes. While other methods based on GPs (e.g., CSId), MI (e.g., tl-CLR) or DTs (e.g., dynGENIE3) had marginally better overall network inference, they do not provide insight into the causal relationships because they only output a positive measure of a causal interaction [28, 33, 40].

Next, we compared the speed of BETS and two other top-performing methods: CSId and Jump3 (Table S3). BETS was the fastest at 4.8 hours while CSId took 9.8 hours and Jump3 took 45 hours. Thus, while BETS had a slightly lower AUPR compared with CSId and Jump3, it was substantially faster.

BETS improved upon Enet using stability selection. To quantify this improvement, we compared three other models: elastic net with lag 1, ridge regression with lag 2, and lasso with lag 2 (Table S4). In each case, the stability selection version outperformed the original version in average AUPR and AUROC. The improvement in average AUPR ranged between 0.016 and 0.03 (+20% to +31%), while the improvement in average AUROC ranged between 0.012 and 0.04 (+1.8% to +6.1%). Hence, our stability selection procedure leads to improved performance for multiple versions of VAR.

We also found that stability selection performance is robust to the number of bootstrap samples (Table S5). Decreasing the number of bootstrap samples from $1,000$ to 100 caused minor decreases of -0.004 in AUPR and -0.008 in AUROC, within the standard deviation across the networks. It also resulted in a 10-fold decrease in memory usage and 3-fold decrease in run time, due to a constant-time hyperparameter search. If users face computational constraints, we recommend that they use 100 bootstrap samples for nearly equivalent performance.

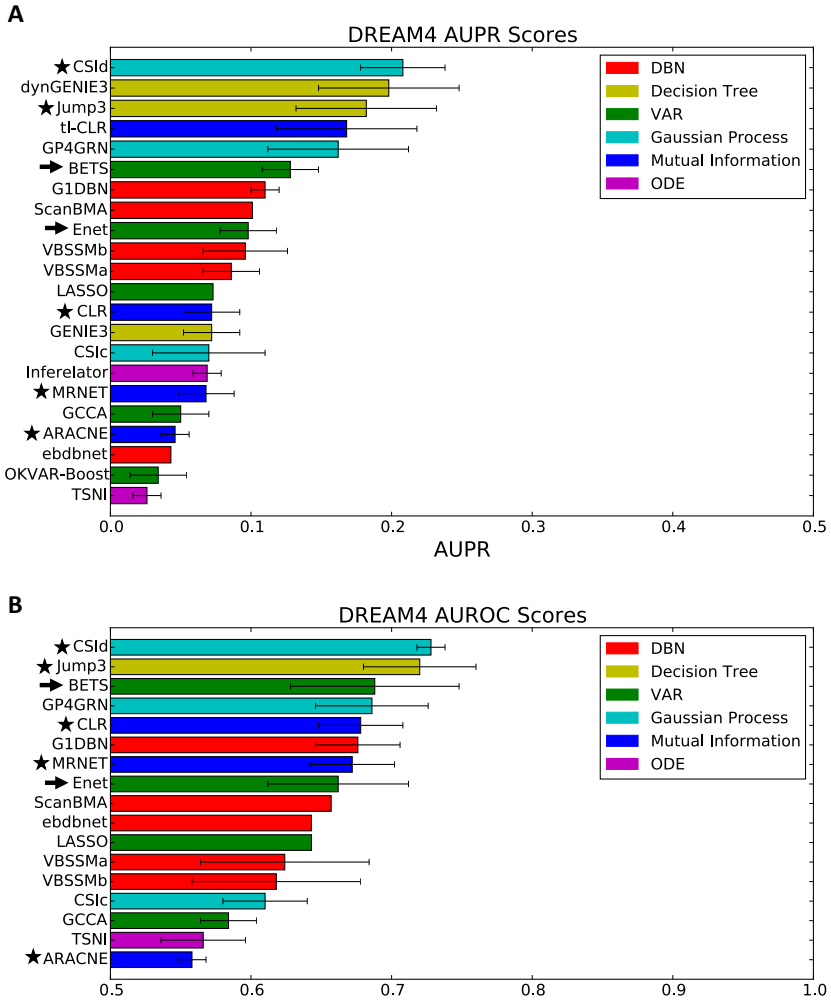

Figure 2: **Algorithm performance on the DREAM Community Benchmark. A) AUPR scores from 22 methods**, averaged across the five DREAM networks. **B) AUROC scores from 17 methods**, averaged across the five DREAM networks. Arrows indicate our methods. Stars indicate methods that we ran in-house; results were consistent with reported results. The bars reach one standard deviation from the average as calculated across the five DREAM networks; no bar indicates the standard deviation is not reported. See also Tables S1, S2, S3, S4, and S5.

3.3. Application to gene transcription response to glucocorticoids.

To infer the causal relationships in the GC response network, we analyzed RNA-seq data collected from the human adenocarcinoma and lung model cell line, A549. This consisted of two data sets. In an *original exposure* data set, cells were exposed to the synthetic GC dexamethasone (dex) for 0, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 10, and 12 hours [6]. In an *unperturbed* data set, the cells were first exposed to dex for 12 hours, after which the media was replaced and dex removed, and then measurements were taken at the same intervals 0, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 10, and 12 hours. BETS was fit jointly over the two data sets. In total there were 7 technical replicates (4 from *original exposure* and 3 from *unperturbed*). A single VAR was fit on 70

samples: Each of the 7 replicates had 10 samples, because using a lag 2 VAR model turns 12 time points into 10 samples.

We applied BETS to the GC-mediated expression responses to infer a causal network (Figure 3A). Edges with selection frequency (frequency of appearance among bootstrap networks) at least 0.097 were declared significant (FDR ≤ 0.2; Figure 3B). The network contained $2,768$ nodes representing distinct genes and $31,945$ directed edges (0.4% of possible edges). Of these, 466 genes were causes (had an outward directed edge) and all $2,768$ genes were effects (had an incoming directed edge). The out-degree distribution was heavy-tailed and skewed right (Figure 3C) while the in-degree distribution was lighter-tailed and more symmetric (Figure 3D). The network's edge in-degree had a heavier left tail and lighter right tail than a normal distribution (Figure 3E). This suggests that causal genes were relatively rare (only 1/6th of network genes were causes) and a fifth of those only affected a single gene, whereas genes that were effects tended to have multiple causes. The network was inferred efficiently due to parallelization across genes, taking six days in real time and 292 days in CPU time to perform 5.5 million elastic net model fits.

To study the network with respect to the glucocorticoid system, we annotated specific genes as transcription factor (TF), immune-related, and metabolism-related [47, 48, 49, 50]. First, we inspected enrichment of each category among the causal genes (Figure 3F). At FDR ≤ 0.05, we found enrichment for TFs among causes; there were 226 causal TFs, representing 8.2% of the $2,768$ input genes. 62 of these TFs were causal, representing 13% of all causal genes (odds ratio (OR) $= 2.0$, Fisher's exact test (FET) adjusted $p \leq 2.9 \times 10^{-5}$). Similarly, we found an enrichment among immune-related genes as causes: of 109 immune genes, representing 3.9% of the input genes, 39 of these were causes, representing 8.4% of all causal genes (OR $= 2.9$, FET adjusted $p \leq 2.5 \times 10^{-6}$). There was no enrichment among metabolism-related genes: there were 120 metabolism genes, representing 4.3% of input genes; 19 of these metabolism genes were causes, representing 4.1% of all causes (OR $= 0.93$, FET adjusted $p \leq 0.66$).

To study the interactions among gene classes inferred by our network, we quantified enrichment for edges between each of the four gene classes – immune, metabolic, TF, and other gene types (*any*) (Figure 3G, Table S6). We found enrichment of 12 of the 16 possible edge types (FDR ≤ 0.05). The network was enriched for edges from any causal genes to immune genes; causal TFs to any genes, TFs, and immune genes; causal immune genes to any genes, TFs, immune genes, and metabolic genes; and causal metabolic genes to any genes, TFs, immune genes, and metabolic genes. This suggests that our network is enriched for causal TFs, immune genes, and metabolic genes.

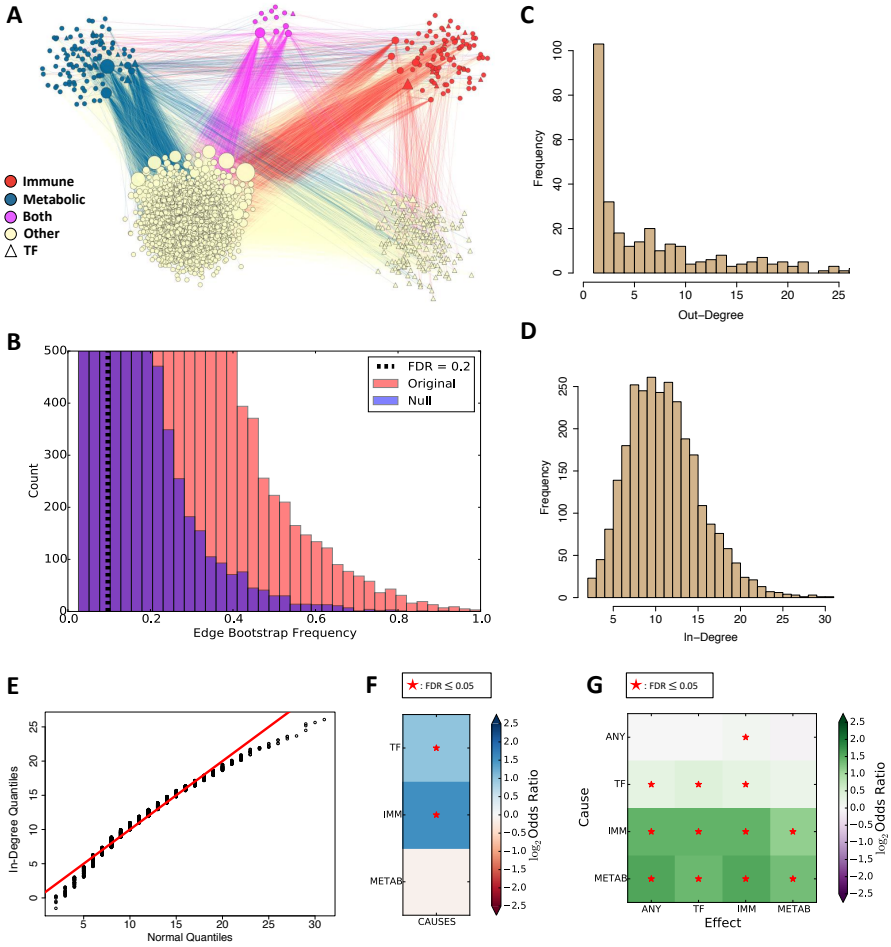

Figure 3: **Causal network inferred from glucocorticoid receptor data. A) Causal network clustered by gene type.** Edge color indicates the type of the causal gene: red edge indicates an immune causal edge, blue edge indicates a metabolic causal edge, purple edge indicates an immune and metabolic causal edge, and tan edge indicates an other causal edge. **B) Significance thresholding for edges**, based on null distribution of selection frequencies. **C) Out-degree distribution of network.** For clarity, several high out-degree values with low frequencies are not shown. **D) In-degree distribution of network. E) Quantile-quantile (Q-Q) plot of in-degree distribution against normal quantiles.** The in-degrees have a heavier left tail and lighter right tail than the normal distribution. **F) Enrichment of gene classes among network causal genes,** measured by odds ratio. **G) Enrichment of edge classes among network edges,** measured by odds ratio. See also Table S6.

Our network identified known biological interactions between genes with immune, metabolic, and TF roles; we highlighted 16 of the gene pairs with experimentally validated interactions (Figure 4, Table S7). *SOCS1* and *SOCS3* bind *IRS2* and promote its degradation, leading to reduced insulin signalling [51, 52]; furthermore, *SOCS1* represses *IL-4*-induced *IRS2* signalling [53]. *NR4A1* heterodimerizes with *RXRA* to activate it to promote gene expression under vitamin A signaling [54]; *NR4A1* also inhibits *p*300-induced *RXRA* acetylation [55]. Eleven of the 16 edges had the correct interaction direction; the five that were

reversed are *TNFAIP3* → *IRAK2*, *SOCS3* → *HIVEP1*, *ATF3* → *MDM2*, *E2F1* → *CDH1*, and *FOS* → *EGFR*. These results suggest that BETS infers biologically meaningful relationships, but transcriptional data, absent other assays on protein abundance and cellular dynamics, are often underpowered to resolve the direction of the edge.

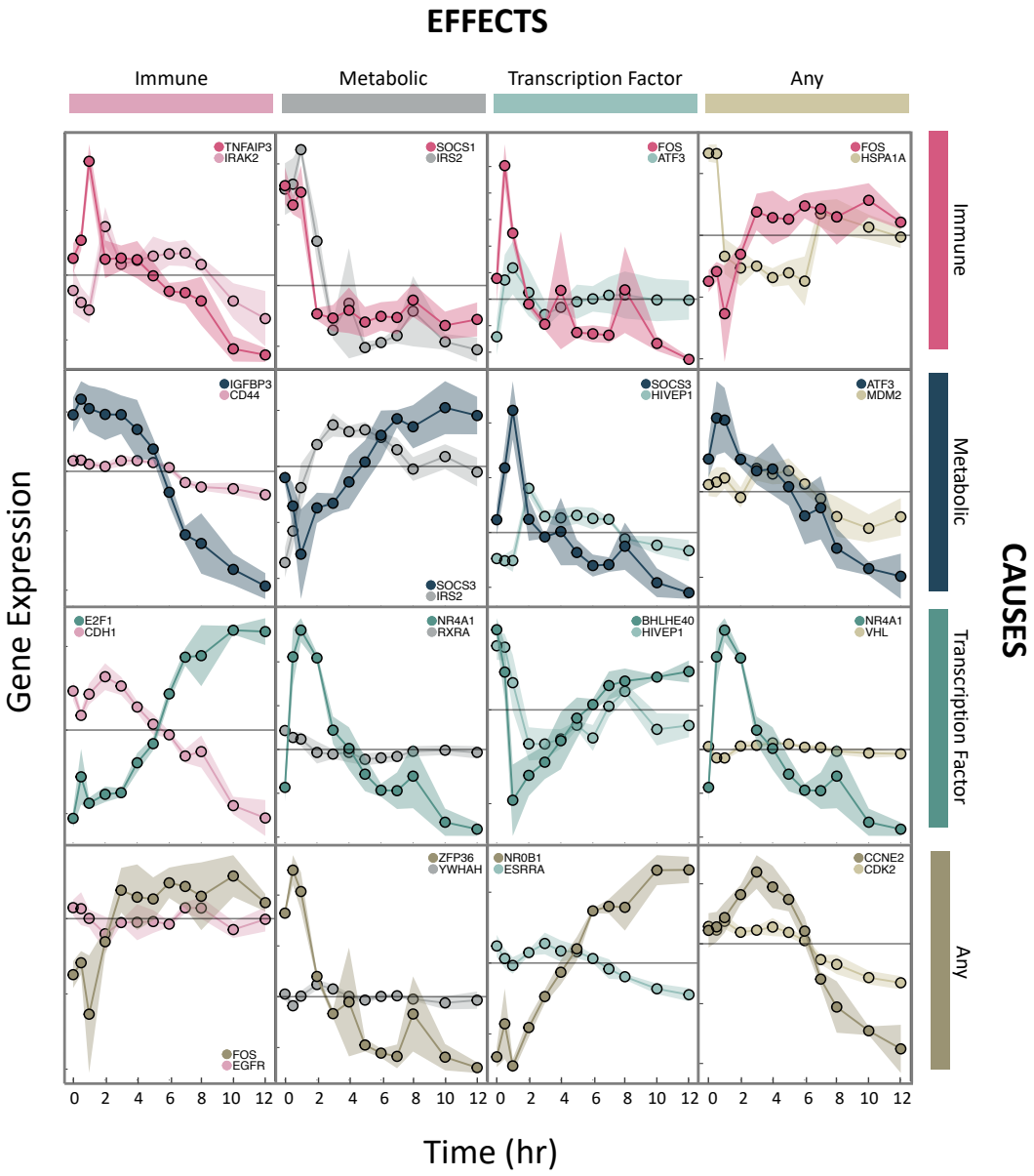

Figure 4: **Time series profiles of experimentally validated causal interactions across gene classes.** For each gene pair, their profiles were from either the *original exposure* data set or the *unperturbed* data set. Because these are conditional Granger-causal relationships, and the effects of the complete set of covariates were not subtracted from the effect gene values, the relationship of the gene pair may be vague. Colors encode gene classes: pink shows immune genes, dark blue/gray shows metabolic genes, teal shows TFs, and brown/tan shows other genes. Darker colors show causal genes and lighter colors show effect genes. The grey line marks zero-centered expression. Each y-axis tick indicates 0.1 unit of ln(TPM) where TPM is Transcripts Per kilobase Million. See also Table S7.

3.4. Validation of inferred network on overexpression data.

We asked whether our inferred network edges validated on overexpression versions of the same experimental system, in which each of ten TFs was separately overexpressed over the same 12 hours of observations. Specifically, we assessed the concordance between inferred network edges $g' \to g$ and their coefficient in the overexpression data set under a VAR model (STAR Methods).

We first evaluated how well network edges replicated on individual overexpression data sets. We performed linear regression of a one-hot encoding of the original network's edge sign (i.e., positive versus no edge or negative sign; negative versus positive or no edge) as the predictor against the VAR model edge coefficients estimated from each of the overexpression time series as the response (Figure 5A-B, STAR Methods). Of the ten data sets, 9 showed enriched positive effect sizes among positive edges at FDR ≤ 0.2 (*CEBPB*, *CEBPD*, *FOSL2*, *FOXO1*, *FOXO3*, *KLF6*, *KLF9*, *KLF15*, *OCT4*; Figure 5A). Three data sets showed enriched negative effect sizes among negative edges (*OCT4*, *TFCP2L1*, *CEBPD*) and four showed enriched positive effect sizes among negative edges (*CEBPB*, *FOSL2*, *KLF9*, *KLF15*; Figure 5B). Taken together, the positive edges inferred by BETS validate on the overexpression data, but the negative edges do not, indicating repressive effects may have inconsistent signs or feedback loops.

Next, we checked whether the 123 inferred causal edges from the TF *TFCP2L1* validated in the *TFCP2L1* overexpression data set (there were only about 10 causal edges from each of the other 9 TFs). We regressed the original network's edge sign (+1 for positive edges, 0 for no edge, and −1 for negative edge) as the predictor against the overexpression VAR model edge coefficients as the response (Figure 5C). We found a positive relationship between the edge sign and overexpression coefficient (slope 0.17, two-sided t-test $p \leq 5 \times 10^{-5}$). This shows that causal edges from *TFCP2L1* are enriched for matched effect directions in the *TFCP2L1* overexpression data.

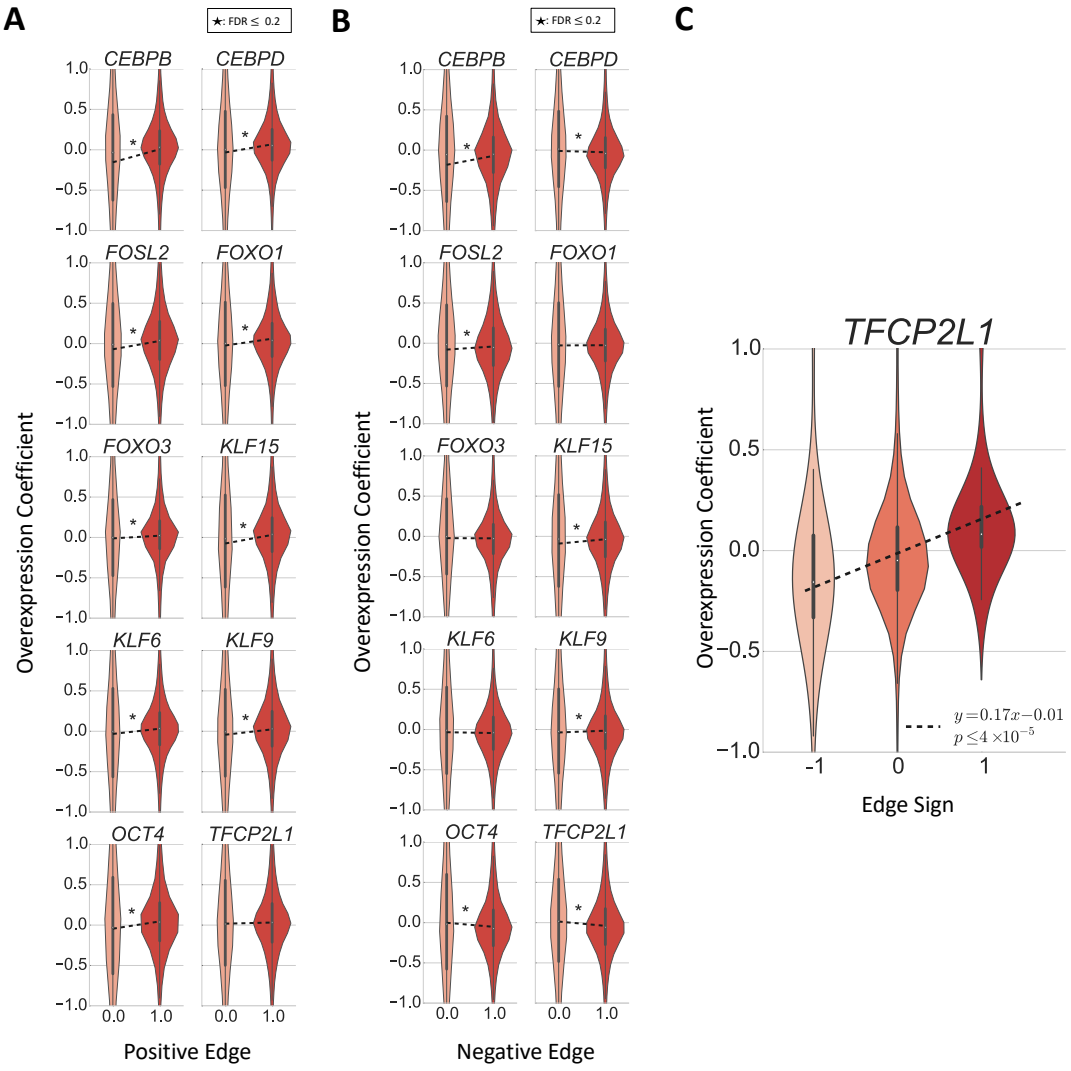

Figure 5: **Validation of inferred network on overexpression data. A-B) Regression of one-hot encoding of positive (negative for B) edges as the predictor against the VAR model edge coefficient from the overexpression data as the response.** A 1 indicates that an edge had a significant positive (in A) or negative (in B) coefficient in the original inferred network (FDR ≤ 0.2). **C) For the 123 causal edges from *TFCP2L1*, regression of edge sign as the predictor against the VAR model edge coefficient from *TFCP2L1* overexpression data as the response.**

3.5. Validation of network edges through lung trans-eQTLs.

We validated our network edges on an expression quantitative trait-loci (eQTL) study. A single nucleotide polymorphism (SNP) S is an eQTL for a gene g' if it is associated with g''s expression level within a population. Given a true causal edge $g' \rightarrow g$, if a SNP S is a local (cis-) eQTL for g', S might also be a distal (trans-) eQTL for g [56]. We used gene expression levels in primary lung tissue ($n = 278$) from the Genotype Tissue Expression (GTEx) project v6p [10]. We observed an enrichment of low trans-eQTL

association p-values from the directed network compared to shuffling the SNP labels (Figure 6A-B). This suggests our network captures more valid causal effects than expected by chance.

We next inspected specific associations and their corresponding edges. We found 341 trans-eQTL pairs in lung samples corresponding with 130 network edges (q-value FDR ≤ 0.2). There are more trans-eQTLs than edges because there are multiple cis-eQTLs for some causal genes g'. The 341 trans-eQTLs greatly improved upon the 2 identified in the GTEx v6 trans-eQTL study [57], demonstrating the utility of transcriptional time series for prioritizing promising associations. The top trans-associations were rs2302178-*CLDN1* (q-value FDR ≤ 0.095, extended from the cis-association rs2302178-*HS3ST6*), rs590429-*ADAMTS* (q-value FDR ≤ 0.11, extended from the cis-association rs2302178-*OLR1*), and rs2072783-*CLIP2* (q-value FDR ≤ 0.11, extended from the cis-association rs2302178-*GMPR*) (Figure 6C).

We searched for validated associations between immune-related genes, metabolic-related genes, and TFs. One association was *OLR1* $\rightarrow$ *ITGAV*, where we see that the known association between SNP rs4329754 and *OLR1* extends to an association between the same SNP and effect gene *ITGAV* (q-value FDR ≤ 0.13) [10]. *OLR1* plays key roles in immunity and metabolism [58, 59]. It is associated with metabolic syndrome [60] and atherosclerosis [60], and modulates inflammatory and humoral immune responses [61, 62]. Meanwhile, *ITGAV* plays a key role in the motility of $CD4^+$ T cells during inflammation [63].

Another association was between the TF *SNAI2* and gene *PTPN6*, where we find that the known association between SNP rs56800165 and *SNAI2* extends to an association between the same SNP rs56800165 and effect gene *PTPN6* (q-value FDR ≤ 0.17) [10]. *SNAI2* is a direct target of the glucocorticoid receptor *GR* to regulate cell migration in breast cancer [64], while *PTPN6* is involved in glucose homeostasis via negatively regulation of insulin signalling [65]. *PTPN6* is also associated with inflammatory phenotypes in multiple diseases [66, 67]. Finally, both *SNAI2* and *PTPN6* are involved in the cell-cell adherens junctions pathway, as *SNAI2* represses transcription of cadherin, while *PTPN6* positively regulates the cadherin-catenin complex [68]. Thus, for several eQTL-validated edges for gene pairs, we find that the genes are involved in related biological processes, but further experimentation is required to confirm direct interactions.

Finally, as A549 cells are models for lung tissue [69], we quantified enrichment of validated edges in lung compared to enrichment in four other tissues: subcutaneous adipose ($n = 298$), transformed fibroblasts ($n = 272$), tibial artery ($n = 285$), and thyroid ($n = 278$). We validated 341 unique network edges across the five tissues (FDR ≤ 0.2). 130 edges validated for lung, 4 for subcutaneous adipose, 125 for transformed fibroblasts, 3 for tibial artery, and 82 for thyroid tissues. More network edges validated in primary lung than in other tissues, suggesting that A549 cells most closely match lung samples among GTEx tissues; this is consistent with their tissue of tumor origin.

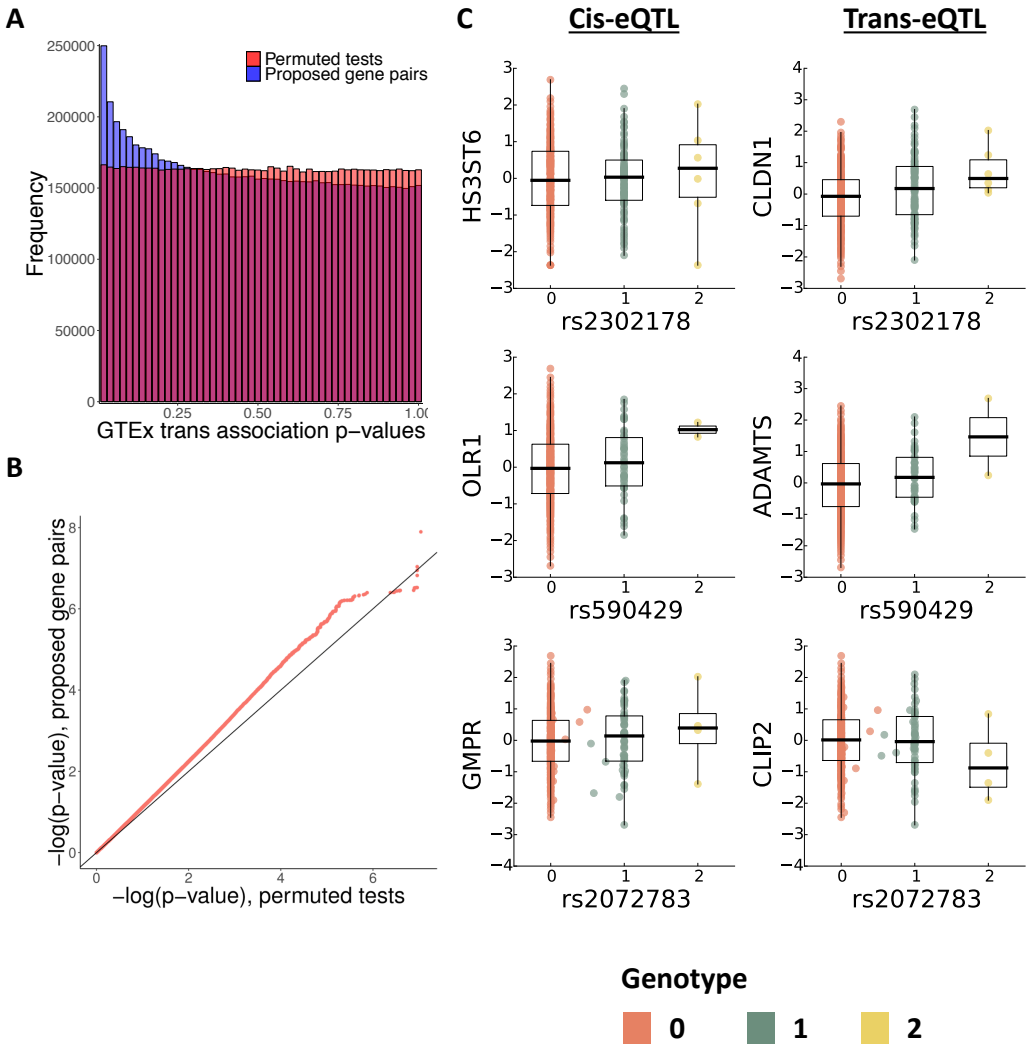

Figure 6: **Network edge validation using known cis- elements from GTEx v6 lung cis-eQTLs. A) Enrichment of trans associations in lung among p-values from edges inferred by BETS** compared to p-values from permutations. **B) Q-Q plot of validated edges** shows signal enrichment in lung samples when compared to signals from four other tissues in the GTEx v6 study. **C) SNPs associated with inferred gene pairs**. Genotype-phenotype plots corresponding to the cis-effect (left column), correlation in the GTEx v6 data between cause (y-axis) and effect (x-axis) gene pairs (right column).

4. Discussion

We described an approach, BETS, to build directed networks using short time series observations of high-dimensional transcriptional data. BETS combined ideas from elastic net regression, graphical lasso, stability selection and VAR models to infer Granger causality relationships in high dimensional transcriptional time

series data. Our method achieved competitive performance on the DREAM4 100-Gene Network Inference Challenge, ranking 6th out of 22 methods in AUPR and 3rd out of 17 methods in AUROC; it was also faster than several methods with similar or better performance and infers effect size and sign, unlike the other top performing methods. Stability selection resulted in consistent improvement to VAR models across different hyperparameter settings.

Next, we applied BETS to time series RNA-seq data from human A549 cells exposed to glucocorticoids and identified a directed network of $31,945$ edges (FDR ≤ 0.2), capturing the causal relationships among genes after exposure to GCs. In our network, we found enrichment of immune genes and TFs among causal genes. We also found enrichment of causal edges from TFs, immune genes, and metabolic genes. We validated our network first in ten overexpression data sets, replicating positive edges with overexpression effects. Validating network edges by searching for trans-eQTLs in GTEx, we found an enrichment of associations with genetic variants across network edges. Finally, we discovered 341 trans-eQTLs, dramatically improving from the GTEx trans-eQTL study without filtering tests for association [57].

While BETS has demonstrated effective inference of causal relations, there are interesting future directions to explore. All methods that infer networks from transcriptional time series face several difficulties. Transcript levels are sometimes an imperfect proxy for protein levels, especially when transcript dynamics are changing [11, 70]; the scarcity of time point samples causes statistical challenges for inferring millions of possible causal interactions between genes, let alone non-additive interactions among causes [3, 71, 72]; transcriptional data do not capture the complete regulatory context including chromatin structure and epigenetic regulations [11]; transcriptional relationships are often nonstationary: the relationship may change over time due to responses from the environment [4, 5]; and inferred networks are often sensitive to the choice of preprocessing and parameter choices [73]. Single cell data also implicitly include transcriptional time series information when pseudotime is inferred, making ideas from Granger causality exceptionally relevant. Finally, experimental followup is key to establishing causality; BETS can only generate promising, interpretable hypotheses. Indeed, by discovering hundreds of more trans-eQTLs than the GTEx study (a 170-fold increase) [57], BETS demonstrates its potential to prioritize biologically meaningful associations.

Author Contributions

TER and BEE designed and funded the study. LKH coordinated all genomic data production. LKH and SML collected RNA-seq data. BD, JL, BJ and BEE developed the methods and validation approaches. BD and JL applied these to data. ICM, TER, BJ, BD, JL, and AB analyzed the data. BJ and JL performed validation in the GTEx data. BD, JL, and BEE drafted the manuscript, and all authors contributed to revision.

This work was funded by the following grants: CZI AWD1005664, CZI AWD1005667, NIH R01 HL133218, NIH U01 HG007900, a Sloan Faculty Fellowship, and an NSF CAREER 1750729.

Acknowledgments

The authors would like to thank Gregory Darnell, Derek Aguiar, Ariel Gewirtz, Allison Chaney, Isabella Grabski, Cristina Anastase, and Genna Gliner for helpful discussion, feedback, and generosity in running cluster jobs; and Jian Peng for productive discussion and helpful comments.

The authors gratefully acknowledge that this work was performed using the Princeton Research Computing resources sponsored by the Princeton Institute for Computational Science and Engineering (PICSciE) at Princeton University.

5. STAR Methods

5.1. Method details

Bootstrap Elastic net regression from Time Series (BETS). Bootstrap Elastic net regression from Time Series (BETS) is a vector-autoregressive approach to causal inference from gene expression time series data. It is based on the principle of Granger causality [27]: a gene g' Granger-causes another gene g if previous information from gene g' improves our current predictions of gene g, beyond using previous information of other genes.

BETS first preprocesses the data. BETS fits an elastic net vector autoregression model to handle the high dimensionality of the time series, inferring a network (Figure 1A). It infers one network for each of $1,000$ bootstrapped samples of the original data set and computes each edge's selection frequency: its frequency of appearance among the bootstrapped networks (Figure 1B) [3]. Finally, BETS includes an edge in the network using the selection frequencies (Figure 1B). Our baseline comparison, Enet, only preprocesses the data and fits an elastic net vector autoregression model from the original data (Figure 1A; Section 3.2).

Preprocessing temporal time series data. For a **gene temporal profile** (i.e., one gene's expression values across time for a single replicate), we used zero-mean unstandardized normalization, which centers each gene temporal profile to have mean zero across time. Because gene temporal profile ranges from staying almost constant to having drastic fluctuations, BETS uses this approach because a unit-variance normalization would over-represent the weak causal effects of genes with lower variability.

Vector autoregression model. Let G be the set of all genes in the data, let $p = |G|$ be the number of genes, and let g be a gene. Let $\neg g$ be G with g removed. Let there be T time points total, and let $t \in \{1, 2, \ldots, T\}$ be a single time point. Let there be R replicates of the gene expression time series.

Let $X^g_{t,r}$ be the expression of gene g at time t for replicate r. Let $X^g_t = [X^g_{t,1}, X^g_{t,2}, \ldots, X^g_{t,R}]^T$ be the $R \times 1$ vector of gene expression levels of gene g across R replicates at time t. The rest of the paper does not mention replicates for simplicity, but here we discuss replicates for completeness.

Let g' be the gene we are testing to be causal for gene g and let ℓ refer to the time lag of the causal edge $g' \to g$. Let L be the maximum lag. In BETS, $L = 2$.

We model each gene g as

$$X^g_t = \sum_{\ell=1}^{L} \alpha^g_\ell X^g_{t-\ell} + \sum_{\ell=1}^{L} \sum_{g' \in \neg g} \beta^{g',g}_\ell X^{g'}_{t-\ell} + \epsilon_t, \tag{2}$$

where $\epsilon_t \sim \mathcal{N}(0, 1)$. In other words, the expression of each gene g is modelled as a linear function of its and other genes' L previous expression values, under independent Gaussian noise. α^g_ℓ represents the (scalar) effect size of gene g's ℓth previous value, $X^g_{t-\ell}$, on its current value, X^g_t. $\beta^{g',g}_\ell$ represents the (scalar) effect size of the ℓth previous value of gene $g' \neq g$, $X^{g'}_{t-\ell}$, on gene g's current value, X^g_t. Equation 2 requires that $t > \ell$ for the ℓth previous value, $X^g_{t-\ell}$, to exist.

To demonstrate how our model is fit in practice, we reformulate Equation 2 using matrix notation. Each row represents one time point for one replicate. There are $T - L$ time points with $t > L$ and R replicates, so there are $R(T - L)$ samples, or rows, in total. Let $N = R(T - L)$.

Define $\mathbf{X}_t^g$, an $N \times 1$ vector, as:

$$\mathbf{X}_t^g = \begin{bmatrix} X_{L+1,1}^g \\ \vdots \\ X_{L+1,R}^g \\ X_{L+2,1}^g \\ \vdots \\ X_{L+2,R}^g \\ \vdots \\ \vdots \\ X_{T,R}^g \end{bmatrix}. \tag{3}$$

We can similarly write $\mathbf{X}_{t-\ell}^g$, which is $\mathbf{X}_t^g$ with each entry replaced by its ℓth previous value. Define $\mathbb{X}_{t-\ell}^g$, a $N \times L$ matrix consisting of the first L previous vectors $\mathbf{X}_{t-\ell}^g$, i.e., for ℓ ranging in $\{1, \ldots, L\}$.

$$\mathbb{X}_{t-\ell}^g = [\mathbf{X}_{t-1}^g \ldots \mathbf{X}_{t-L}^g]. \tag{4}$$

Let $\boldsymbol{\alpha}_\ell^g$ be a $L \times 1$ vector of the L lagged coefficients.

$$\boldsymbol{\alpha}_\ell^g = \begin{bmatrix} \alpha_1^g \\ \vdots \\ \alpha_L^g \end{bmatrix}. \tag{5}$$

Next, let us formulate Equation 2 involving the genes g' in matrix notation. Let $\mathbb{X}_{t-\ell}^{\neg g}$ be a $N \times L(|G|-1)$ predictor matrix of the vectors $\mathbf{X}_{t-\ell}^{g'}$, for $g' \neq g$ and $\ell \in \{1, \ldots L\}$. Note the number of columns is $L(|G|-1)$, because there are L previous time points $\ell \in \{1, \ldots, L\}$, and for each ℓ, there are $|G|-1$ genes $g' \neq g$, giving $|G|-1$ vectors: $\mathbf{X}_{t-\ell}^{g_1'}, \ldots, \mathbf{X}_{t-\ell}^{g_{|G|-1}'}$.

$$\mathbb{X}_{t-\ell}^{\neg g} = \left[\mathbf{X}_{t-1}^{g_1'} \ldots \mathbf{X}_{t-1}^{g_{|G|-1}'} \mathbf{X}_{t-2}^{g_1'} \ldots \mathbf{X}_{t-2}^{g_{|G|-1}'} \ldots\ldots \mathbf{X}_{t-L}^{g_{|G|-1}'}\right]. \tag{6}$$

Let $\boldsymbol{\beta}_\ell^{\cdot,g}$ be a $L(|G|-1) \times 1$ vector of the causal coefficients $\boldsymbol{\beta}_\ell^{g',g}$ where $g' \neq g$.

$$\boldsymbol{\beta}_\ell^{\cdot,g} = \begin{bmatrix} \beta_1^{g_1',g} \\ \vdots \\ \beta_1^{g_{|G|-1}',g} \\ \beta_2^{g_1',g} \\ \vdots \\ \beta_2^{g_{|G|-1}',g} \\ \vdots \\ \vdots \\ \beta_L^{g_{|G|-1}',g} \end{bmatrix}. \tag{7}$$

We then fit the model:

$$\mathbf{X}_t^g = \mathbb{X}_{t-\ell}^g \boldsymbol{\alpha}_\ell^g + \mathbb{X}_{t-\ell}^{\neg g} \boldsymbol{\beta}_\ell^{\cdot,g} + \epsilon_t, \tag{8}$$

where $\boldsymbol{\epsilon}_t$ is a $N \times 1$ vector with each element $\epsilon_{t,n} \sim \mathcal{N}(0,1)$. To write in the most compact form, we can write

$$\mathbb{X}_{t-\ell}^G = [\mathbb{X}_{t-\ell}^g \mathbb{X}_{t-\ell}^{\neg g}], \quad \bar{\beta}^g = \begin{bmatrix} \boldsymbol{\alpha}_\ell^g \\ \boldsymbol{\beta}_\ell^{\cdot,g} \end{bmatrix}. \tag{9}$$

Note that $\mathbb{X}_{t-\ell}^G$ is a $N \times L|G|$ matrix and $\bar{\beta}^g$ is a $L|G| \times 1$ vector. Thus the final matrix formulation of Equation 2 is:

$$\mathbf{X}_t^g = \mathbb{X}_{t-\ell}^G \bar{\beta}^g + \boldsymbol{\epsilon}_t. \tag{10}$$

Elastic net penalty. Because of the large number of predictors as compared to the small number of samples, we use the elastic net penalty, which is a generalization of both ridge and lasso penalties. The elastic net fits the following objective:

$$\hat{\beta}^g_{\text{ELASTIC NET}} = \underset{\bar{\beta}^g \in \mathbb{R}^{L|G|}}{\arg\min} \|\mathbf{X}_t^g - \mathbb{X}_{t-\ell}^G \bar{\beta}^g\|_2^2 + \lambda(a\|\bar{\beta}^g\|_1 + (1-a)\|\bar{\beta}^g\|_2^2). \tag{11}$$

Here $\|\cdot\|_1$ represents the ℓ_1-norm and $\|\cdot\|_2$ represents the ℓ_2-norm.

For the elastic net, we used the following ranges of hyperparameter values: $\lambda \in \{10^{-4}, 10^{-3}, \ldots, 1\}, a \in \{0.1, 0.3, \ldots, 0.9\}$. For lasso, we used $\lambda \in \{10^{-5}, \ldots, 1\}$. For ridge, when we used $\{10^{-5}, \ldots, 1\}$, we found that the the optimal value selected in some cases was the maximum value of $\lambda = 1$. We thus expanded the range to $\{10^{-5}, \ldots, 10^6\}$ to ensure that we were not missing better hyperparameters at larger values. At this point, the optimal λ was found to be 100.

Hyperparameter tuning. Hyperparameters were selected using leave-one-out cross-validation (LOOCV). The hyperparameter (or pair of hyperparameters, for elastic net) that minimizes the mean-squared error on the held-out datapoints is selected. More specifically, we first fix a hyperparameter (λ, a). Then, for a given gene g and row index i, extract the i-the row of $\mathbf{X}_t^g$ and $\mathbb{X}_{t-\ell}^G$. Refer to this extracted validation set as $(\mathbf{X}_t^g)_i$ (target) and $(\mathbb{X}_{t-\ell}^g)_i$ (predictors). The remaining data is the training set, $(\mathbf{X}_t^g)_{-i}$ (target) and $(\mathbb{X}_{t-\ell}^G)_{-i}$ (predictors).

First, let $\hat{\beta}^g_{(\lambda,a),i}$ be the $\hat{\beta}^g_{\text{ELASTIC NET}}$ that is fit from the training set.

$$\hat{\beta}^g_{(\lambda,a),i} = \underset{\bar{\beta}^g \in \mathbb{R}^{L|G|}}{\arg\min} \|(\mathbf{X}_t^g)_{-i} - (\mathbb{X}_{t-\ell}^G)_{-i} \bar{\beta}^g\|_2^2 + \lambda(a\|\bar{\beta}^g\|_1 + (1-a)\|\bar{\beta}^g\|_2^2). \tag{12}$$

We then compute prediction error on the validation set, $\|(\mathbf{X}_t^g)_i - (\mathbb{X}_{t-\ell}^G)_i \hat{\beta}^g_{(\lambda,a),i}\|_2^2)$. We repeat the fit $\hat{\beta}^g_{(\lambda,a),i}$ and error for every row index i of $\mathbf{X}_t^g$ and for every gene g. The mean held-out cross-validation error for (λ, a) is:

$$MSE(\lambda, a) = \sum_{g \in G} \sum_{i=1}^{N} \frac{1}{N} \|(\mathbf{X}_t^g)_i - (\mathbb{X}_{t-\ell}^G)_i \hat{\beta}^g_{(\lambda,a),i}\|_2^2. \tag{13}$$

The (λ, a) that minimizes the error in Equation 13 is selected.

Permuted coefficients. We evaluate the significance of any given edge $g' \to g$ through permutation. In detail, we remove the time dependency between g' and g via permutations of individual gene temporal profiles over time.

We first generate a single permuted data set $\tilde{X}_t^g$. For each gene, we independently shuffle the temporal profile of each gene $g \in \{1, \ldots, |G|\}$ across time (Figure 1A). This is done separately for distinct replicates.

We wish to model the hypothesis of no causal relations from any gene $g' \in \neg g$, upon a given effect gene g. We use the unpermuted values of the effect gene X_t^g and the permuted values of all other causal genes $g' \in \neg g$, as $\tilde{X}_t^{\neg g}$. The effect gene g remains unpermuted, as we do not consider self-regulatory loops.

Permutation-based causal coefficients $\tilde{\beta}_\ell^{g',g}$ are then fit as

$$X_t^g = \sum_{\ell=1}^{L} \alpha_\ell^g X_{t-\ell}^g + \sum_{\ell=1}^{L} \sum_{g' \in \neg g} \tilde{\beta}_\ell^{g',g} \tilde{X}_{t-\ell}^{g'} + \epsilon_t. \tag{14}$$

We use these coefficients to perform FDR calibration.

Edge FDR. The result of the elastic net VAR model is a complete network whose edges are weighted according to the estimated regression coefficients.

For each lag $\ell \in \{1, \ldots, L\}$ and effect gene g, we control the edge FDR at ≤ 0.05 by finding the threshold T_ℓ^g such that

$$\frac{\sum_{g' \in \neg g} \mathbb{1}\{|\tilde{\beta}_\ell^{g',g}| > T_\ell^g\}}{\sum_{g' \in \neg g} \mathbb{1}\{|\tilde{\beta}_\ell^{g',g}| > T_\ell^g\} + \sum_{g' \in \neg g} \mathbb{1}\{|\beta_\ell^{g',g}| > T_\ell^g\}} \leq 0.05. \tag{15}$$

For each gene pair (g', g), $g' \in \neg g$, a directed edge $g' \to g$ exists if for at least one of the lags $\ell \in \{1, \ldots, L\}$, $|\beta_\ell^{g',g}| > T_\ell^g$.

Stability selection. Stability selection is used to ensure the robustness of BETS to small sample size. Stability selection is a method for high-dimensional graph estimation that uses bootstrap samples [74]. While the authors prove finite sample control for the family-wise error rate (FWER), we are interested in controlling the false discovery rate (FDR).

First, we draw $B = 1,000$ bootstrap samples, where each sample consists of $N = R(T - L)$ rows drawn with replacement from $\mathbb{X}_{t-\ell}^G$, the predictors, and $\mathbf{X}_t^g$, the target (Equation 10).

For each bootstrap sample, we infer a network using BETS. Each edge $g' \to g$'s selection frequency, $\pi_{g',g}$ (the frequency of $g' \to g$ among the bootstrap networks) is computed. (Figure 1 B).

Stability FDR. To determine the appropriate cutoff for the selection frequency of each edge ($\pi_{g',g}$), we generate a null distribution of selection frequencies using permutations. First, we generate a second permuted data set in which we again independently shuffle the temporal profile of each gene $g \in \{1, \ldots, |G|\}$ across time. This is done separately for distinct replicates. We run the selection frequency procedure on this permuted data set to get the null selection frequency of each edge, $\tilde{\pi}_{g',g}$.

We control the stability FDR at 0.2 by finding the threshold T_b such that

$$\frac{\sum_{g' \in \neg g} \mathbb{1}\{\tilde{\pi}_{g',g} > T_b\}}{\sum_{g' \in \neg g} \mathbb{1}\{\tilde{\pi}_{g',g} > T_b\} + \sum_{g' \in \neg g} \mathbb{1}\{\pi_{g',g} > T_b\}} \leq 0.2. \tag{16}$$

Because the maximum lag is 2, each edge $g' \to g$ has two possible lags and thus two selection frequencies. The lag with larger absolute value of average coefficient across the $1,000$ networks is considered in both the

permuted and the real empirical distributions. So, if $|\beta_1^{g',g}|$ exceeds $|\beta_2^{g',g}|$, the lag is said to be 1 and the selection frequency $\pi^1_{g',g}$ is used.

Network inference performance metrics. Refer to every network edge inferred by a method as a positive and every missing edge as a negative. Let TP be True Positives, FP be False Positives, TN be True Negatives, and FN be False Negatives. Let TPR be True Positive Rate, (i.e., recall), and FPR be False Positive Rate. Then, we have

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

In the DREAM benchmark, each network inference method is evaluated by comparing the true network (i.e., the network used to generate the synthetic data) with the inferred network at different thresholds for edge inclusion. The two main evaluation metrics are Area Under the Receiver Operating Characteristic curve (AUROC) and Area Under the Precision-Recall curve (AUPR). AUROC plots TPR on the y axis and FPR on the x axis. AUPR plots precision on the y axis and recall on the x axis. When the number of negatives greatly exceeds the number of positives, as with gene networks, which are typically sparse, AUPR is a more relevant metric [75].

5.2. Software

BETS is available for download on Github at `https://github.com/lujonathanh/BETS`. The software is licensed under the terms of the Apache License, version 2.0.

5.3. Data sets and Processing

DREAM Network Inference Challenge. There were 5 data sets in the DREAM4 Network Inference Challenge, each consisting of 10 time series of 21 time points and 100 genes [41, 76]. For the first half of the time series, a "drug perturbation" was applied; this affected about 1/3 of genes. For the second half, the perturbation was removed and the system was allowed to relax back to the wild-type state.

Glucocorticoid gene expression data. We analyzed RNA sequencing data from a set of experiments developed to study glucocorticoid receptors (GRs) in the human adenocarcinoma and lung model cell line, A549 [6]. There was an *original exposure* data set of 4 replicates in which cells were stimulated by the glucocorticoid dexamethasone (dex), and gene expression was profiled at $\{0, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 10, 12\}$ hours of dex stimulation. There was also an *unperturbed* data set of 3 replicates in which cells were exposed to dex for 12 hours, after which the conditioned media was replaced and dex removed. Gene expression was profiled at $\{0, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 10, 12\}$ hours after dex removal. We integrated the *original exposure* and *unperturbed* data into a joint data set with 7 replicates.

We selected $2,768$ genes for analysis, which had average expression > 2 TPM and were differentially expressed in the *original exposure data*. A gene was called differentially expressed if its expression at any time point differed from its expression at time 0, ascertained by running edgeR (FDR ≤ 0.05) [6]. We added

NR3C1, which encodes the glucocorticoid receptor (GR). *NR3C1* was not found to be differentially expressed at FDR ≤ 0.05.

After genes were selected, gene expression Transcripts Per kilobase Million (TPM) were log-normalized and corrected for surrogate variables using SVAseq [77]. Each gene's temporal profile was centered to have mean zero across time. In the *original exposure data*, all replicates besides replicate 1 had a measurement for each time point. Replicate 1 was missing time points 5 and 6 hrs, so we imputed these values using a linear interpolation from time points 4 and 7 hrs in the log-transformed, surrogate-corrected space.

Overexpression transcriptional time series data. There were 10 *overexpression* data sets, in which each of the transcription factors $CEBPB$, $CEBPD$, $FOSL2$, $FOXO1$, $FOXO3$, $KLF6$, $KLF9$, $KLF15$, $POU5F1$, and $TFCP2L1$ was separately overexpressed across 12 hours of dex stimulation. Each overexpression data set had three replicates; gene expression was profiled after $\{0, 1, 4, 8, 12\}$ hours of dex stimulation. The same 2768 genes were selected and the same normalization and SVAseq correction as earlier was performed.

5.4. Application of methods to the data

DREAM benchmarking. We ran the methods BETS, Enet, CSId [40], Jump3 [32], CLR [23], MRNET [24], and ARACNE [25] on the DREAM challenge. In BETS, inferred edges were ranked by their selection frequency for calculating AUPR and AUROC. In Enet, edges were ranked by the absolute value of their coefficient. The Python3 version of CSId was run after obtaining it from correspondence with Dr. Penfold.

Jump3 required setting the "systematic noise" and "observational noise" parameters. We used Dr. Huynh-Thu's settings on the DREAM challenge, with systematic noise at $1e-4$ and observational noise at 0.01 times the value of the gene's expression. ARACNE, MRNET, and CLR were run using the minet R library. BETS, Enet, CSId, and Jump3 were run on a single node without parallelization. The node had 28 cores, 128 GB of memory, and 2.4 GHz processor speed. ARACNE, MRNET, and CLR were run on a 4 GB RAM, Intel Core i5 1.3 GHz laptop.

Network analysis: Gene annotations. We considered genes with three possible labels: immune system, metabolism, or transcription factor. Immune genes were labeled as such using two sources. The first source is the Gene Ontology (GO) annotation "Immune" (*GO:0002376*) [47]. We applied this label when the evidence codes were EXP, IDA, IGI, IMP, IPI, IC, TAS. The second source is the Gene Ontology Consortium's curated, ranked list of immune-related genes based on multiple databases and experimental evidence [49]. For the GO annotation, we selected all genes with score ≥ 7. This resulted in 616 immune genes overall, and 109 immune genes in our list of 2768 genes.

Metabolic genes were called using two sources. The first source is the GO annotation "carbohydrate metabolic process" *GO:0005975* [47]. We applied this label when the evidence codes were EXP, IDA, IGI, IMP, IPI, IC, TAS. The second source is the Gene Set Enrichment Analysis (GSEA)-curated list of metabolic-related genes [48]. We searched only among those with experimental evidence: the Canonical, KEGG, BIOCARTA, and Reactome pathways. We used the following four search queries: "gluconeogenesis OR (glucose AND metabolism) OR glycolysis," "lipid AND metabolism," "Diabetes," "Obesity." This resulted in 544 metabolic genes overall, of which 120 were in our gene list. 65 genes were both immune and metabolic overall; 12 of these were in our gene list.

Transcription factors (TFs) were called using the Bioguo database of human TFs [50]. There were 1463 TFs overall, of which 226 were present in our gene list.

Experimental interactions. We created a list of experimentally validated interactions from the BIOGRID Homo sapiens Protein-Protein Interactions database [78] and the STRING database [79]. Proteins were mapped to genes using BioMart from Ensembl 94 [80]. Among genes in our gene list, there were $17,990$ BIOGRID interactions and $13,148$ STRING interactions.

Validation on overexpression data. The overexpression data had four time points with 1 to 4 hour time gaps, unlike the original 12 time points with 0.5 to 2 hour time gaps. On the overexpression data, we used a VAR model that regressed each effect gene's expression level on its previous expression level and the causal gene's previous expression level, assuming normal noise $\epsilon_t \sim \mathcal{N}(0,1)$:

$$X_t^g = c^g X_{t-1}^g + d^{g',g} X_{t-1}^{g'} + \epsilon_t. \tag{17}$$

No regularization was included, and ordinary least squares was used to fit the equation. The expression $X_{t-1}^{g'}$ of a causal gene g' is fit as a single predictor without the other expression. Lag 1, not 2, is used due to the larger time gaps.

Validation on lung trans-eQTLs in GTEx v6. Trans-eQTLs were discovered using the Genotype Tissue Expression (GTEx) v6 data [10, 57]. First, we mapped our genes from hg38 to hg19. For every edge $g' \to g$, we tested the set of single nucleotide polymorphisms (SNPs) within 20 kilobases of g' for trans-eQTL association with g [81]. Specifically, we computed the p-value for linear association of each SNP with the corresponding effect gene g using MatrixEQTL [82]. A null distribution was generated by taking every edge $g' \to g$, permuting the effect gene g's expression values, and repeating the linear association test. FDR over test statistics was calculated using q-value [83]. Because not every causal gene g' had a cis-eQTL, only $26,839$ edges (84% of the original $31,945$ edges) were tested.

6. Supplemental Information

6.1. Network inference methods

Several methods have been developed to estimate directed graphs of genes from transcriptional time series data (Figure S1). Broadly, these methods estimate directed networks in which the directed edges between nodes—representing genes—indicate a cause-effect relationship between those genes, such that perturbing the expression levels of the causal gene would lead to changes in expression of the effect gene [20].

Let G be the set of all genes and g be a single gene. Let $\neg g$ be G with g removed. Let there be T time points total, and let t be a single time point ranging from $\{1, 2, \ldots, T\}$. Let X_t^g be the expression of gene g at time t. Let ϵ_t be the residual noise at time t. Let : denote sequencing through values, for example $X_{1:T}^g$ would denote all the values X_1^g through X_T^g. Let $pa(X_t^g)$ refer to the causal parents of gene g at time t in dynamic Bayesian Networks. For example, $pa(X_t^g)$ may include $X_{t-\ell}^{g'}$. Let g' be the gene we are testing to be causal for gene g. Let ℓ be the time lag of the causal interaction. We are testing the existence of the edge $g' \to g$ at lag ℓ.

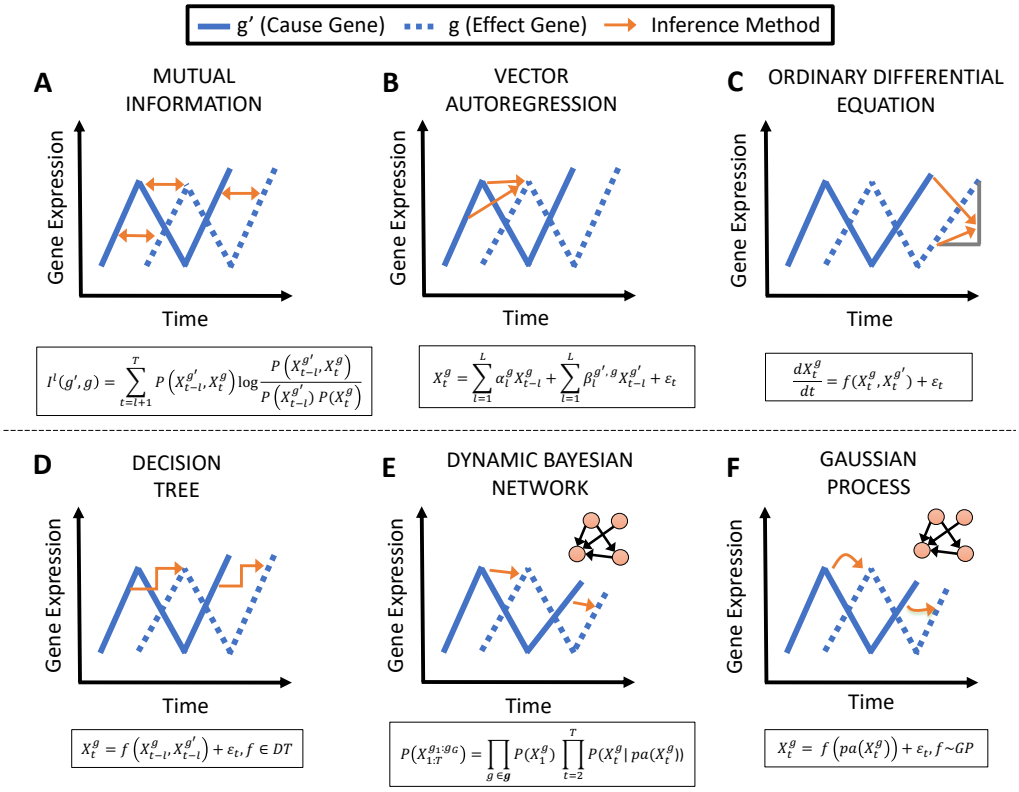

Figure S1: Related to Figure 1. **Overview of gene regulatory network inference methods.** Panels show each inference method applied to a cause gene g' (blue, solid) and an effect gene g (blue, dotted). A) Mutual information is computed between the cause and effect. B) The effect's expression is fit as an autoregression from the cause's past expression. C) The effect's expression is fit as a differential equation from the cause's current expression. D) The effect's expression is fit as a decision tree function of the cause's past expression. E) The space of dynamic causal networks is searched, with linear relationships between cause and effect. F) The space of dynamic causal networks is searched, with nonlinear relationships between cause and effect.

6.1.1. Mutual information

Mutual information (MI) methods assess the MI between the expression of g' at the ℓ-th previous time point and the expression of g at the current time point (Figure S1A) [21, 22, 23, 24, 25]:

$$I^{\ell}(g', g) = \sum_{t=l+1}^{T} P(X_{t-\ell}^{g'}, X_t^{g}) \log \frac{P(X_{t-\ell}^{g'}, X_t^{g})}{P(X_{t-\ell}^{g'}) P(X_t^{g})}. \tag{18}$$

A causal edge $g' \rightarrow g$ is included if $I^{\ell}(g', g)$ exceeds a threshold. MI methods have the advantage of being simple and fast. However, they do not give insight into the sign of two genes' relationship (i.e., activation or repression) because MI is an unsigned metric [24, 26].

6.1.2. *Granger causality*

Granger causality methods determine if including the expression of g' at the previous time point improves our ability to predict the expression of g at the current time point above using the expression of g at the previous time point [27]. A common way to implement a Granger causality approach uses a vector autoregression (VAR) model, which usually assumes a linear relationship between all genes' previous expression and g's current expression. (Figure S1B) [84].

$$X_t^g = \sum_{\ell=1}^{L} \alpha_\ell^g X_{t-\ell}^g + \sum_{g' \in \neg g} \sum_{\ell=1}^{L} \beta_\ell^{g',g} X_{t-\ell}^{g'} + \epsilon_t. \tag{19}$$

A causal edge $g' \to g$ is included in the network if $\beta_\ell^{g',g}$ is significantly different from 0 for some ℓ. While older VAR analyses did not fit the causal predictors simultaneously [84, 85, 86], newer analyses fit them simultaneously, using regularization techniques such as lasso [13, 87] or ridge regression [88] to handle the high dimensionality of genome-wide sequencing assays. Nonlinear, kernel-valued functions have also been used to implement ideas in Granger causality [89].

6.1.3. *Ordinary differential equations*

Ordinary differential equations (ODEs) fit the derivative of the expression of g as a function of all genes' expression at a single time point (Figure S1C) [11, 28, 29]:

$$\frac{dX_t^g}{dt} = f(X_t^g, X_t^{g'_1}, \ldots X_t^{g'_{G-1}}) + \epsilon_t. \tag{20}$$

Although complex dynamics are often nonlinear, ODE methods typically assume linearity, as small sample sizes make it challenging to infer the parameters of nonlinear functions. A causal edge $g' \to g$ is included in the network if g' has a significant coefficient in the ODE.

These methods are often combined with additional methods such as spline interpolation and piecewise linear functions to improve performance [28, 29].

6.1.4. *Decision trees*

Decision trees (DT) are a type of nonparametric function based on partitioning the data [30, 31]. DT methods fall either under VAR or ODE methods. Either the DTs fit the expression of g at the current time as a function of all genes' expression at the previous time point (VAR), or they fit the derivative of the expression of g as a function of all genes' expression at a single time point (ODE) (Figure S1D) [32, 33].

$$X_t^g = f(X_{t-1:t-L}^g, X_{t-1:t-L}^{g'_1:g'_{|G|-1}}) + \epsilon_t, \quad f \in DT. \tag{21}$$

$$\frac{dX_t^g}{dt} = f(X_t^g, X_t^{g'_1}, \ldots X_t^{g'_{|G|-1}}) + \epsilon_t, \quad f \in DT. \tag{22}$$

A causal edge $g' \to g$ is included in the network when an importance score for g'— typically, the reduction in variance of g from including g' as a predictor— exceeds some threshold. One limitation of DT methods is that they only produce a ranking of edges, without specifying the sign of the relationship between the genes [32, 33].

6.1.5. Dynamic Bayesian networks

Dynamic Bayesian networks (DBNs) search the space of possible directed acyclic graphs between previous and current expression levels and identify the network structure with the highest posterior probability of each edge given the data (Figure S1E) [34, 35, 36, 37, 38]. DBNs typically assume a linear relationship between previous expression values and current expression values. A causal edge $g' \rightarrow g$ is included in the network when its marginal posterior probability of existence exceeds some threshold. The joint probability of all the genes' expression across the time points $1:T$ factorizes as:

$$P(X_{1:T}^{g_1:g_{|G|}}) = \prod_{g \in G} P(X_1^g) \prod_{t=2}^{T} P(X_t^g | pa(X_t^g)).$$

Each gene's expression has a linear relationship with its parents:

$$X_t^g = f(pa(X_t^g)) + \epsilon_t, \quad f \in \text{LINEAR}.$$

While DBNs have been shown to be effective on smaller data sets [90], they scale poorly due to the super-exponential growth of possible causal graph structures [56, 88]. Even after limiting the number of possible parents per gene to two, this results in cubic scaling of the search space. One exception is ScanBMA, which uses a pruning method based on Occam's window to limit the search space and gain a speedup [36].

6.1.6. Gaussian process

The Gaussian process (GP) is a distribution over continuous, nonlinear functions. GPs are often used in the context of nonlinear DBNs, where GP regression is used to model a nonlinear relationship between previous expression levels and current expression levels (Figure S1F) [39, 40]. A causal edge $g' \rightarrow g$ is included in the network based on its posterior probability of existence, i.e., the sum of the posterior probabilities of those networks that contain the edge. Each gene's expression has a nonlinear relationship with its parents:

$$X_t^g = f(pa(X_t^g)) + \epsilon_t, \quad f \sim GP.$$

By allowing nonlinear relationships between genes, GPs have proven highly effective. However, like DBN, they perform a search over causal graphs, and therefore suffer from the same scalability issues [39, 40].

6.2. Validation of inferred network on overexpression data

Our analyses regressed network edge signs as predictors against the VAR model edge coefficients from the overexpression data as response (Figure 5). We sought to assess the strength of these associations across the 10 data sets, compared against shuffled edges. We compared the effect sizes of all 10 regressions of positive edge one-hot encodings on the overexpression coefficients with the effect sizes estimated similarly after shuffling the edge labels; we did the same for negative edges. At FDR ≤ 0.2, there was a substantial enrichment of effect sizes of positive edges among the original network (Common Language Effect Size (CLES) $= 0.93$, two-sided Mann-Whitney U-test (MWU) adjusted $p \leq 0.0026$); there was no enrichment of effect sizes for negative edges in the original network (CLES $= 0.55$, two-sided MWU adjusted $p \leq 0.73$). Thus, the positive edges inferred by BETS validate on the overexpression data, but the negative edges do not, indicating repressive effects may have inconsistent signs.

6.3. Supplementary Tables

Algorithm	Method Type	AUPR (Average)	AUPR (STD)	Network 1 AUPR	Network 2 AUPR	Network 3 AUPR	Network 4 AUPR	Network 5 AUPR
ebdbnet	DBN	0.043						
G1DBN	DBN	0.11	0.01	0.11	0.1	0.13	0.1	0.11
ScanBMA	DBN	0.101						
VBSSMa	DBN	0.086	0.02	0.08	0.05	0.11	0.1	0.09
VBSSMb	DBN	0.096	0.03	0.09	0.06	0.12	0.12	0.09
dynGENIE3	DT	0.198	0.05	0.22	0.14	0.25	0.22	0.16
GENIE3	DT	0.072	0.02	0.05	0.06	0.1	0.06	0.09
Jump3	DT	0.182	0.05	0.26	0.11	0.19	0.17	0.18
CSIc	GP	0.07	0.04	0.13	0.03	0.07	0.07	0.05
CSId	GP	0.208	0.03	0.26	0.17	0.22	0.2	0.19
GP4GRN	GP	0.162	0.05	0.22	0.1	0.16	0.21	0.12
ARACNE	MI	0.046	0.01	0.03	0.04	0.06	0.04	0.06
CLR	MI	0.072	0.02	0.05	0.06	0.11	0.06	0.08
MRNET	MI	0.068	0.02	0.04	0.06	0.1	0.06	0.08
tl-CLR	MI	0.168	0.05	0.18	0.11	0.24	0.15	0.16
Inferelator	ODE	0.0688	0.01	0.063	0.071	0.075	0.073	0.062
TSNI	ODE	0.026	0.01	0.02	0.03	0.03	0.02	0.03
BETS	VAR	0.128	0.02	0.16	0.1	0.13	0.14	0.11
Enet	VAR	0.098	0.02	0.12	0.08	0.1	0.11	0.08
GCCA	VAR	0.05	0.02	0.04	0.04	0.07	0.07	0.03
LASSO	VAR	0.073						
OKVAR-Boost	VAR	0.034	0.02	0.05	0.05	0.03	0.02	0.02

Table S1: Related to Figure 2. **DREAM4 100-gene Network Inference Results, AUPR**. DBN is Dynamic Bayesian Network, DT is Decision Tree, GP is Gaussian Process, MI is Mutual Information, ODE is Ordinary Differential Equation, VAR is Vector Autoregression. The references that reported ebdbnet, ScanBMA, and LASSO did not provide AUPR values for individual networks. Algorithms that were run in-house were ARACNE, BETS, CLR, CSId, Enet, Jump3 and MRNET. Where reported literature values were available, they were consistent with these values. Values for CSIc, G1DBN, GCCA, GP4GRN, TSNI, VBSSMa and VBSSMb were taken from [45]. Values for ebdnet, LASSO and ScanBMA, were taken from [36]. Values for dynGENIE3, GENIE3, OKVAR-Boost and tl-CLR were taken from [33]. Value for Inferelator and Jump3 were taken from [32].

Algorithm	Method Type	Average AUROC	STD AUROC	Network 1 AUROC	Network 2 AUROC	Network 3 AUROC	Network 4 AUROC	Network 5 AUROC
ebdbnet	DBN	0.643						
G1DBN	DBN	0.676	0.03	0.68	0.64	0.68	0.66	0.72
ScanBMA	DBN	0.657						
VBSSMa	DBN	0.624	0.06	0.59	0.56	0.59	0.67	0.71
VBSSMb	DBN	0.618	0.06	0.56	0.57	0.62	0.64	0.7
Jump3	DT	0.72	0.04	0.77	0.67	0.74	0.68	0.74
CSIc	GP	0.61	0.03	0.65	0.56	0.63	0.61	0.6
CSId	GP	0.728	0.01	0.74	0.71	0.72	0.74	0.73
GP4GRN	GP	0.686	0.04	0.72	0.62	0.7	0.7	0.69
ARACNE	MI	0.558	0.01	0.56	0.54	0.56	0.55	0.58
CLR	MI	0.678	0.03	0.7	0.63	0.71	0.67	0.68
MRNET	MI	0.672	0.03	0.68	0.63	0.71	0.66	0.68
TSNI	ODE	0.566	0.03	0.55	0.55	0.6	0.54	0.59
BETS	VAR	0.688	0.06	0.78	0.65	0.64	0.7	0.67
Enet	VAR	0.662	0.05	0.73	0.62	0.62	0.67	0.67
GCCA	VAR	0.584	0.02	0.6	0.57	0.6	0.58	0.57
LASSO	VAR	0.643						

Table S2: Related to Figure 2. **DREAM4 100-gene Network Inference Results, AUROC**. DBN is Dynamic Bayesian Network, DT is Decision Tree, GP is GP, MI is MI, ODE is Ordinary Differential Equation, VAR is Vector Autoregression. The references that reported ebdbnet, ScanBMA, and LASSO did not provide AUROC values for individual networks. Algorithms that were run in-house were ARACNE, BETS, CLR, CSId, Enet, Jump3 and MRNET. Values for CSIc, G1DBN, GCCA, GP4GRN, TSNI, VBSSMa and VBSSMb were taken from [45]. Values for ebdnet, LASSO and ScanBMA, were taken from [36].

Method	AUPR		AUROC		Time (hours)	Literature
	In-House	Literature	In-House	Literature	In-House	Reference
CSId	0.208	0.234	0.728	0.712	9.2	[45]
Jump3	0.182	0.187	0.72		45	[32]
BETS	**0.128**		**0.688**		**4.8**	
Enet	**0.098**		**0.662**		**1.2**	
CLR	0.072	0.123	0.678	0.699	0.0000089	[36]
MRNET	0.068	0.13	0.672	0.701	0.000011	[36]
ARACNE	0.046	0.106	0.558	0.589	0.000010	[36]

Table S3: Related to Figure 2. **Results of In-House Algorithms on DREAM4 100-gene Network Inference.** AUPR and AUROC indcate average AUPR and AUROC over the 5 networks, respectively. BETS and Enet are bolded to indicate that they are our own developed methods, based on vector autoregression. CSId is a Gaussian process method [40]. Jump3 is a decision tree method [32]. CLR [23], MRNET [91], and ARACNE [25] are mutual information methods. See Supplemental Information for discussion of differences when compared to literature-reported performance for ARACNE, MRNET, and CLR.

Normalization	Lag	Penalty	Coefficient AUROC	Bootstrap AUROC	Coefficient AUPR	Bootstrap AUPR
0mean	1	Elastic Net	0.674 (0.05)	0.686 (0.05)	0.112 (0.03)	0.14 (0.03)
0mean	2	Elastic Net	0.662 (0.05)	0.688 (0.06)	0.098 (0.02)	0.128 (0.02)
0mean	2	Lasso	0.652 (0.05)	0.692 (0.06)	0.14 (0.04)	0.162 (0.05)
0mean	2	Ridge	0.642 (0.04)	0.66 (0.05)	0.08 (0.03)	0.096 (0.03)

Table S4: Related to Figure 2. **Improvement on DREAM4 100-gene Network Inference from Bootstrap.** For each AUROC or AUPR column, the average is the listed value and the standard deviation is listed in parentheses. "Coefficient" denotes the result when ranking edges by their fitted coefficient, as in the original method. "Bootstrap" denotes the results when ranking edges by the frequency by which they appear in the bootstrap networks.

Bootstrap Samples	AUROC	AUPR	Time (hr)	Memory (GB)
100	0.68 (0.05)	0.124 (0.02)	1.6	1.6
1000	**0.688 (0.06)**	**0.128 (0.02)**	**4.8**	**15.6**

Table S5: Related to Figure 2. **Dependency of BETS performance on Bootstrap Samples.** DREAM results reported for running BETS on both 100 and 1000 bootstrap samples. All values in the columns are averages and the parenthetical values as standard deviations across the 5 DREAM4 Networks. The 1000 samples row is bolded because 1000 samples are the default settings. These use zero-mean normalization, lag 2, and the elastic net penalty.

Cause Type	Effect Type	Odds Ratio	Raw p-value	Adjusted p-value	Significant at FDR 0.05
Any	Any	1	1.00E+00	1.00E+00	No
Any	TF	1	5.69E-01	6.51E-01	No
Any	Imm	1.06	2.73E-02	3.64E-02	Yes
Any	Metab	0.96	9.16E-01	9.77E-01	No
TF	Any	1.22	3.51E-25	9.36E-25	Yes
TF	TF	1.32	4.43E-06	7.09E-06	Yes
TF	Imm	1.21	2.24E-02	3.25E-02	Yes
TF	Metab	1.14	7.54E-02	9.28E-02	No
Imm	Any	2.69	0.00E+00	0.00E+00	Yes
Imm	TF	2.71	2.21E-45	1.18E-44	Yes
Imm	Imm	2.71	5.98E-23	1.37E-22	Yes
Imm	Metab	2.18	5.65E-14	1.01E-13	Yes
Metab	Any	2.96	0.00E+00	0.00E+00	Yes
Metab	TF	2.58	1.55E-43	6.20E-43	Yes
Metab	Imm	2.93	2.45E-30	7.85E-30	Yes
Metab	Metab	2.49	6.21E-22	1.24E-21	Yes

Table S6: Related to Figure 3. **Enrichment of edges between specific gene classes in inferred causal network**. A Fisher's Exact Test was performed, whether the rows of the contingency table were whether or not an edge was of the edge type, and the columns were whether or not the edge was part of the inferred network.

Edge Type	Cause Gene	Effect Gene	Lag	Average Coefficient	Selection Frequency	Shown Data set	Correct Direction?	Citation
I → I	TNFAIP3	IRAK2	1	0.01	0.163	*unperturbed*	N	[92]
I → M	SOCS1	IRS2	1	0.01	0.1	*unperturbed*	Y	[51, 53]
I → T	FOS	ATF3	1	0.1	0.709	*unperturbed*	Y	[93]
I → A	FOS	HSPA1A	2	0.02	0.2	*original*	Y	[94]
M → I	IGFBP3	CD44	1	0.01	0.212	*original*	Y	[95]
M → M	SOCS3	IRS2	2	0.01	0.099	*original*	Y	[52]
M → T	SOCS3	HIVEP1	1	0.02	0.297	*unperturbed*	N	[96]
M → A	ATF3	MDM2	1	0.02	0.46	*original*	N	[97]
T → I	E2F1	CDH1	1	-0.01	0.12	*unperturbed*	N	[98]
T → M	NR4A1	RXRA	2	0.02	0.471	*unperturbed*	Y	[54, 55, 99]
T → T	BHLHE40	HIVEP1	1	0.01	0.152	*original*	Y	[96]
T → A	NR4A1	VHL	2	0.01	0.237	*unperturbed*	Y	[100]
A → I	FOS	EGFR	1	0.004	0.109	*original*	N	[101]
A → M	ZFP36	YWHAH	2	0.004	0.131	*unperturbed*	Y	[102]
A → T	NR0B1	ESRRA	1	-0.004	0.102	*unperturbed*	Y	[103]
A → A	CCNE2	CDK2	1	0.04	0.636	*original*	Y	[104, 105, 106, 107]

Table S7: Related to Figure 4. **Gene Pair information from Figure 4.** Shown Data Set indicates whether the gene temporal profiles in Figure 4 are taken from the *original exposure data* or *unperturbed data*. The edge type indicates the gene class of the causal and effect gene; for example, I → M indicates an edge from an Immune causal gene to a Metabolic effect gene. I = Immune; M = Metabolic; T = Transcription Factor; A = Any gene.

[1] Z. Bar-Joseph, A. Gitter, I. Simon, Studying and modelling dynamic biological processes using time-series gene expression data, Nature Reviews Genetics 13 (2012) 552–564.

[2] J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, M. West, Bayesian factor regression models in the "large p, small n" paradigm, Bayesian Statistics 7 (2003) 733–742.

[3] P. Bühlmann, M. Kalisch, L. Meier, High-dimensional statistics with a view toward applications in biology, Annual Review of Statistics and Its Application 1 (2014) 255–278.

[4] P. Mas, Circadian clock function in arabidopsis thaliana: time beyond transcription, Trends in cell biology 18 (2008) 273–281.

[5] J. W. Robinson, A. J. Hartemink, Learning non-stationary dynamic bayesian networks, Journal of Machine Learning Research 11 (2010) 3647–3680.

[6] I. C. McDowell, A. Barrera, A. M. D'Ippolito, C. M. Vockley, L. K. Hong, S. M. Leichter, L. C. Bartelt, W. H. Majoros, L. Song, A. Safi, et al., Glucocorticoid receptor recruits to enhancers and drives activation by motif-directed binding, Genome Research 28 (2018) 1272–1284.

[7] D. W. Cain, J. A. Cidlowski, Immune regulation by glucocorticoids, Nature Reviews Immunology (2017).

[8] E. B. Geer, J. Islam, C. Buettner, Mechanisms of glucocorticoid-induced insulin resistance: focus on adipose tissue function and lipid metabolism, Endocrinology and Metabolism Clinics of North America 43 (2014) 75–102.

[9] S. J. Spencer, A. Tilbrook, The glucocorticoid contribution to obesity, Stress 14 (2011) 233–246.

[10] GTEx Consortium, et al., Genetic effects on gene expression across human tissues, Nature 550 (2017) 204.

[11] Z.-P. Liu, Reverse Engineering of Genome-wide Gene Regulatory Networks from Gene Expression Data, Current Genomics 16 (2015) 3–22.

[12] R. Opgen-Rhein, K. Strimmer, Learning causal networks from systems biology time course data: an effective model selection procedure for the vector autoregressive process, BMC Bioinformatics 8 (2007) 1.

[13] A. C. Lozano, N. Abe, Y. Liu, S. Rosset, Grouped graphical Granger modeling for gene expression regulatory networks discovery, Bioinformatics 25 (2009) i110–i118.

[14] H. Cho, B. Berger, J. Peng, Reconstructing Causal Biological Networks through Active Learning, PloS One 11 (2016) e0150611.

[15] M. H. Maathuis, M. Kalisch, P. Bühlmann, et al., Estimating high-dimensional intervention effects from observational data, The Annals of Statistics 37 (2009) 3133–3164.

[16] K. P. Murphy, Active Learning of Causal Bayes Net Structure, Technical Report, University of California, Berkeley, 2001.

[17] A. Rau, F. Jaffrézic, G. Nuel, Joint estimation of causal effects from observational and intervention gene expression data, BMC Systems Biology 7 (2013) 1.

[18] A. Hauser, P. Bühlmann, Two optimal strategies for active learning of causal models from interventional data, International Journal of Approximate Reasoning 55 (2014) 926–939.

[19] Y.-B. He, Z. Geng, Active learning of causal networks with intervention experiments and optimal designs, Journal of Machine Learning Research 9 (2008) 2523–2547.

[20] M. Grzegorczyk, An introduction to Gaussian Bayesian networks, Systems Biology in Drug Discovery and Development: Methods and Protocols (2010) 121–147.

[21] A. Madar, A. Greenfield, E. Vanden-Eijnden, R. Bonneau, DREAM3: network inference using dynamic context likelihood of relatedness and the inferelator, PloS One 5 (2010) e9803.

[22] M. Lopes, G. Bontempi, Experimental assessment of static and dynamic algorithms for gene regulation inference from time series expression data, Frontiers in Genetics 4 (2013) 303.

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

[24] P. E. Meyer, K. Kontos, F. Lafitte, G. Bontempi, Information-theoretic inference of large transcriptional regulatory networks, EURASIP Journal on Bioinformatics and Systems Biology 2007 (2007) 8–8.

[25] A. A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. Dalla Favera, A. Califano, ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context, in: BMC Bioinformatics, volume 7, BioMed Central, p. S7.

[26] P. Zoppoli, S. Morganella, M. Ceccarelli, TimeDelay-ARACNE: Reverse engineering of gene networks from time-course data by an information theoretic approach, BMC Bioinformatics 11 (2010) 154.

[27] C. Granger, Testing for causality, Journal of Economic Dynamics and Control 2 (1980) 329 – 352.

[28] M. Bansal, G. D. Gatta, D. Di Bernardo, Inference of gene regulatory networks and compound mode of action from time course gene expression profiles, Bioinformatics 22 (2006) 815–822.

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

[30] J. R. Quinlan, Induction of decision trees, Machine Learning 1 (1986) 81–106.

[31] L. Breiman, Classification and regression trees, Routledge, 2017.

[32] V. A. Huynh-Thu, G. Sanguinetti, Combining tree-based and dynamical systems for the inference of gene regulatory networks, Bioinformatics 31 (2015) 1614–1622.

[33] P. Geurts, et al., dynGENIE3: dynamical GENIE3 for the inference of gene networks from time series expression data, Scientific Reports 8 (2018) 3384.

[34] S. Lèbre, Inferring dynamic genetic networks with low order independencies, Statistical Applications in Genetics and Molecular Biology 8 (2009) 1–38.

[35] A. J. Hartemink, D. K. Gifford, T. S. Jaakkola, R. A. Young, et al., Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks., in: Pacific Symposium on Biocomputing, volume 6, p. 266.

[36] W. C. Young, A. E. Raftery, K. Y. Yeung, Fast Bayesian inference for gene regulatory networks using ScanBMA, BMC Systems Biology 8 (2014) 47.

[37] M. J. Beal, F. Falciani, Z. Ghahramani, C. Rangel, D. L. Wild, A Bayesian approach to reconstructing genetic regulatory networks with hidden factors, Bioinformatics 21 (2004) 349–356.

[38] A. Rau, F. Jaffrézic, J.-L. Foulley, R. W. Doerge, An empirical Bayesian method for estimating biological networks from temporal microarray data, Statistical Applications in Genetics and Molecular Biology 9 (2010).

[39] T. Äijö, H. Lähdesmäki, Learning gene regulatory networks from gene expression measurements using non-parametric molecular kinetics, Bioinformatics 25 (2009) 2937–2944.

[40] C. A. Penfold, A. Shifaz, P. E. Brown, A. Nicholson, D. L. Wild, CSI: a nonparametric Bayesian approach to network inference from multiple perturbed time series gene expression data, Statistical Applications in Genetics and Molecular Biology 14 (2015) 307–310.

[41] D. Marbach, T. Schaffter, C. Mattiussi, D. Floreano, Generating realistic in silico gene networks for performance assessment of reverse engineering methods, Journal of Computational Biology 16 (2009) 229–239.

[42] J. Friedman, T. Hastie, R. Tibshirani, Sparse inverse covariance estimation with the graphical lasso, Biostatistics 9 (2008) 432–441.

[43] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67 (2005) 301–320.

[44] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society. Series B (Methodological) (1996) 267–288.

[45] C. A. Penfold, D. L. Wild, How to infer gene networks from expression profiles, revisited, Interface Focus 1 (2011) 857–870.

[46] A. Irrthum, L. Wehenkel, P. Geurts, et al., Inferring regulatory networks from expression data using tree-based methods, PloS ONE 5 (2010) e12776.

[47] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, et al., Gene Ontology: tool for the unification of biology, Nature Genetics 25 (2000) 25–29.

[48] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles, Proceedings of the National Academy of Sciences 102 (2005) 15545–15550.

[49] T. G. O. Consortium, Gene Ontology Consoritum's Curated List of Immune Genes, `http://wiki.geneontology.org/index.php/Immunology`, 2014. Accessed: 2017-04-22.

[50] H.-M. Zhang, H. Chen, W. Liu, H. Liu, J. Gong, H. Wang, A.-Y. Guo, AnimalTFDB: a comprehensive animal transcription factor database, Nucleic Acids Research 40 (2012) D144–D149.

[51] L. Rui, M. Yuan, D. Frantz, S. Shoelson, M. F. White, SOCS-1 and SOCS-3 block insulin signaling by ubiquitin-mediated degradation of IRS1 and IRS2, Journal of Biological Chemistry 277 (2002) 42394–42398.

[52] V. C. Calegari, M. Alves, P. K. Picardi, R. Y. Inoue, K. G. Franchini, M. J. Saad, L. A. Velloso, Suppressor of cytokine signaling-3 provides a novel interface in the cross-talk between angiotensin II and insulin signaling systems, Endocrinology 146 (2005) 579–588.

[53] S. M. McCormick, N. Gowda, J. X. Fang, N. M. Heller, Suppressor of cytokine signaling (SOCS) 1 regulates IL-4-activated insulin receptor substrate (IRS)-2 tyrosine phosphorylation in monocytes and macrophages via the proteasome, Journal of Biological Chemistry (2016) jbc–M116.

[54] T. Perlmann, L. Jansson, A novel pathway for vitamin A signaling mediated by RXR heterodimerization with NGFI-B and NURR1., Genes & Development 9 (1995) 769–782.

[55] W.-x. Zhao, M. Tian, B.-x. Zhao, G.-d. Li, B. Liu, Y.-y. Zhan, H.-z. Chen, Q. Wu, Orphan receptor TR3 attenuates the p300-induced acetylation of retinoid X receptor-α, Molecular Endocrinology 21 (2007) 2877–2889.

[56] J. Peters, Causality: Lecture Notes, ETH Zurich, ETH Zurich, 2015.

[57] B. Jo, Y. He, B. J. Strober, P. Parsana, F. Aguet, A. A. Brown, S. E. Castel, E. R. Gamazon, A. Gewirtz, G. Gliner, B. Han, A. Z. He, E. Y. Kang, I. C. McDowell, X. Li, P. Mohammadi, C. B. Peterson, G. Quon, A. Saha, A. V. Segre, J. H. Sul, T. J. Sullivan, K. G. Ardlie, C. D. Brown, D. F. Conrad, N. J. Cox, E. T. Dermitzakis, E. Eskin, M. Kellis, T. Lappalainen, C. Sabatti, B. E. Engelhardt, A. Battle, Distant regulatory effects of genetic variation in multiple human tissues, bioRxiv (2016).

[58] P. C. Chui, H.-P. Guan, M. Lehrke, M. A. Lazar, PPARγ regulates adipocyte cholesterol metabolism via oxidized LDL receptor 1, The Journal of Clinical Investigation 115 (2005) 2244–2256.

[59] C. Arslan, B. Bayoglu, C. Tel, M. Cengiz, A. Dirican, K. Besirli, Upregulation of OLR1 and IL17A genes and their association with blood glucose and lipid levels in femoropopliteal artery disease, Experimental and Therapeutic Medicine 13 (2017) 1160–1168.

[60] V. O. Palmieri, B. Coppola, I. Grattagliano, V. Casieri, G. Cardinale, P. Portincasa, G. Palasciano, F. Di Serio, Oxidized LDL receptor 1 gene polymorphism in patients with metabolic syndrome, European Journal of Clinical Investigation 43 (2013) 41–48.

[61] S. Oh, H. Joo, LOX-1 boosts immunity, Oncotarget 6 (2015) 21763.

[62] H. Joo, D. Li, M. Dullaers, T.-W. Kim, D. Duluc, K. Upchurch, Y. Xue, S. Zurawski, R. Le Grand, Y.-J. Liu, et al., C-type lectin-like receptor LOX-1 promotes dendritic cell-mediated class-switched B cell responses, Immunity 41 (2014) 592–604.

[63] M. G. Overstreet, A. Gaylo, B. Angermann, A. Hughson, Y.-m. Hyun, K. Lambert, M. Acharya, A. C. Billroth-Maclurg, A. F. Rosenberg, D. J. Topham, et al., Inflammation-induced effector CD4+ T cell interstitial migration is alpha-v integrin dependent, Nature Immunology 14 (2013) 949.

[64] J. Ling, A. Singhal, Z. P. Lopez-Dee, B. Porreca, T. Sprague, Snai2 is a new target to mediate glucocorticoid signaling on breast cancer cell migration, in: Proceedings of the American Association of Cancer Research Annual Meeting, July 2018, volume 78, AACR, 2018.

[65] M.-J. Dubois, S. Bergeron, H.-J. Kim, L. Dombrowski, M. Perreault, B. Fournès, R. Faure, M. Olivier, N. Beauchemin, G. I. Shulman, et al., The SHP-1 protein tyrosine phosphatase negatively modulates glucose homeostasis, Nature Medicine 12 (2006) 549.

[66] K. W. Eriksen, A. Woetmann, L. Skov, T. Krejsgaard, L. F. Bovin, M. L. Hansen, K. Grønbæk, N. Billestrup, M. H. Nissen, C. Geisler, et al., Deficient SOCS3 and SHP-1 expression in psoriatic T cells, Journal of Investigative Dermatology 130 (2010) 1590–1597.

[67] G. P. Christophi, M. Panos, C. A. Hudson, R. L. Christophi, R. C. Gruber, A. T. Mersich, S. D. Blystone, B. Jubelt, P. T. Massa, Macrophages of multiple sclerosis patients display deficient SHP-1 expression and enhanced inflammatory phenotype, Laboratory Investigation 89 (2009) 742.

[68] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, K. Morishima, KEGG: new perspectives on genomes, pathways, diseases and drugs., Nucleic Acids Res 45 (2017) D353–D361.

[69] M. Lieber, G. Todaro, B. Smith, A. Szakal, W. Nelson-Rees, A continuous tumor-cell line from a human lung carcinoma with properties of type ii alveolar epithelial cells, International journal of cancer 17 (1976) 62–70.

[70] Y. Liu, A. Beyer, R. Aebersold, On the dependency of cellular protein levels on mRNA abundance, Cell 165 (2016) 535–550.

[71] R. De Smet, K. Marchal, Advantages and limitations of current network inference methods, Nature Reviews Microbiology 8 (2010) 717.

[72] D. Marbach, R. J. Prill, T. Schaffter, C. Mattiussi, D. Floreano, G. Stolovitzky, Revealing strengths and weaknesses of methods for gene network inference, Proceedings of the National Academy of Sciences (2010).

[73] S. Uygun, C. Peng, M. D. Lehti-Shiu, R. L. Last, S.-H. Shiu, Utility and limitations of using gene expression data to identify functional associations, PLoS Computational Biology 12 (2016) e1005244.

[74] N. Meinshausen, P. Bühlmann, Stability selection, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72 (2010) 417–473.

[75] J. Davis, M. Goadrich, The relationship between Precision-Recall and ROC curves, in: Proceedings of the 23rd International Conference on Machine Learning, ACM, pp. 233–240.

[76] D. Marbach, T. Schaffter, D. Floreano, R. J. Prill, G. Stolovitsky, The DREAM4 In-silico Network Challenge: Training data, gold standards, and supplementary information, `http://gnw.sourceforge.net/resources/DREAM4%20in%20silico%20challenge.pdf`, 2009.

[77] J. T. Leek, J. D. Storey, Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis, PLOS Genetics 3 (2007) 1–12.

[78] A. Chatr-Aryamontri, R. Oughtred, L. Boucher, J. Rust, C. Chang, N. K. Kolas, L. O'Donnell, S. Oster, C. Theesfeld, A. Sellam, et al., The BioGRID interaction database: 2017 update, Nucleic Acids Research 45 (2017) D369–D379.

[79] D. Szklarczyk, J. H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N. T. Doncheva, A. Roth, P. Bork, et al., The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible, Nucleic Acids Research (2016) gkw937.

[80] D. R. Zerbino, P. Achuthan, W. Akanni, M. R. Amode, D. Barrell, J. Bhai, K. Billis, C. Cummins, A. Gall, C. G. Girón, et al., Ensembl 2018, Nucleic Acids Research 46 (2017) D754–D761.

[81] C. Gao, S. Zhao, I. C. McDowell, C. D. Brown, B. E. Engelhardt, Context-specific and differential gene co-expression networks via Bayesian biclustering models, PLOS Computational Biology 12 (2016) e1004791.

[82] A. A. Shabalin, Matrix eQTL: ultra fast eQTL analysis via large matrix operations, Bioinformatics 28 (2012) 1353–1358.

[83] J. D. Storey, R. Tibshirani, Statistical significance for genomewide studies, Proceedings of the National Academy of Sciences 100 (2003) 9440–9445.

[84] G. H. F. Tam, C. Chang, Y. S. Hung, Application of Granger causality to gene regulatory network discovery, in: Systems Biology (ISB), 2012 IEEE 6th International Conference on, IEEE, pp. 232–239.

[85] N. D. Mukhopadhyay, S. Chatterjee, Causality and pathway search in microarray time series experiment, Bioinformatics 23 (2007) 442.

[86] J. Zhu, Y. Chen, A. S. Leonardson, K. Wang, J. R. Lamb, V. Emilsson, E. E. Schadt, Characterizing Dynamic Changes in the Human Blood Transcriptional Network, PLoS Computational Biology 6 (2010).

[87] A. Shojaie, G. Michailidis, Discovering graphical Granger causality using the truncating lasso penalty, Bioinformatics 26 (2010) i517–i523.

[88] S. Yao, S. Yoo, D. Yu, Prior knowledge driven Granger causality analysis on gene regulatory network discovery, BMC Bioinformatics 16 (2015) 273.

[89] N. Lim, Y. Şenbabaoğlu, G. Michailidis, F. dAlché Buc, OKVAR-Boost: a novel boosting algorithm to infer nonlinear dynamics and interactions in gene regulatory networks, Bioinformatics 29 (2013) 1416–1423.

[90] C. Zou, J. Feng, Granger causality vs. dynamic Bayesian network inference: a comparative study, BMC Bioinformatics 10 (2009) 122.

[91] P. E. Meyer, F. Lafitte, G. Bontempi, minet: AR/Bioconductor package for inferring large transcriptional networks using mutual information, BMC Bioinformatics 9 (2008) 461.

[92] Y. Xiong, F. Qiu, W. Piao, C. Song, L. M. Wahl, A. E. Medvedev, Endotoxin tolerance impairs IL-1 receptor-associated kinase (IRAK) 4 and TGF-β-activated kinase 1 activation, K63-linked polyubiquitination and assembly of IRAK1, TNF receptor-associated factor 6, and IκB kinase γ and increases A20 expression, Journal of Biological Chemistry 286 (2011) 7905–7916.

[93] A. W. Reinke, J. Baek, O. Ashenberg, A. E. Keating, Networks of bZIP protein-protein interactions diversified over a billion years of evolution, Science 340 (2013) 730–734.

[94] E. Miyamoto-Sato, S. Fujimori, M. Ishizaka, N. Hirai, K. Masuoka, R. Saito, Y. Ozawa, K. Hino, T. Washio, M. Tomita, et al., A comprehensive resource of interacting protein regions for refining human transcription factor networks, PloS ONE 5 (2010) e9289.

[95] M. Natsuizaka, H. Kinugasa, S. Kagawa, K. A. Whelan, S. Naganuma, H. Subramanian, S. Chang, K. J. Nakagawa, N. L. Rustgi, Y. Kita, et al., IGFBP3 promotes esophageal cancer growth by suppressing oxidative stress in hypoxic tumor microenvironment, American Journal of Cancer Research 4 (2014) 29.

[96] J. Wang, K. Huo, L. Ma, L. Tang, D. Li, X. Huang, Y. Yuan, C. Li, W. Wang, W. Guan, et al., Toward an understanding of the protein interaction network of the human liver, Molecular Systems Biology 7 (2011) 536.

[97] P. Mo, H. Wang, H. Lu, D. D. Boyd, C. Yan, MDM2 mediates ubiquitination and degradation of activating transcription factor 3, Journal of Biological Chemistry (2010) jbc–M110.

[98] M. J. Peart, M. V. Poyurovsky, E. M. Kass, M. Urist, E. Verschuren, M. K. Summers, P. K. Jackson, C. Prives, APC/CCdc20 targets E2F1 for degradation in prometaphase, Cell Cycle 9 (2010) 3956–3964.

[99] B. Lin, S. K. Kolluri, F. Lin, W. Liu, Y.-H. Han, X. Cao, M. I. Dawson, J. C. Reed, X.-k. Zhang, Conversion of Bcl-2 from protector to killer by interaction with nuclear orphan receptor Nur77/TR3, Cell 116 (2004) 527–540.

[100] B.-Y. Kim, H. Kim, E.-J. Cho, H.-D. Youn, Nur77 upregulates HIF-α by inhibiting pVHL-mediated degradation, Experimental & Molecular Medicine 40 (2008) 71.

[101] T.-C. Chen, K.-T. Lin, C.-H. Chen, S.-A. Lee, P.-Y. Lee, Y.-W. Liu, Y.-L. Kuo, F.-S. Wang, J.-M. Lai, C.-Y. F. Huang, Using an in situ proximity ligation assay to systematically profile endogenous protein–protein interactions in a pathway network, Journal of Proteome Research 13 (2014) 5339–5346.

[102] B. A. Johnson, J. R. Stehn, M. B. Yaffe, T. K. Blackwell, Cytoplasmic localization of Tristetraprolin involves 14-3-3-dependent and–independent mechanisms, Journal of Biological Chemistry (2002).

[103] X. Li, W. Wang, J. Wang, A. Malovannaya, Y. Xi, W. Li, R. Guerra, D. H. Hawke, J. Qin, J. Chen, Proteomic analyses reveal distinct chromatin-associated and soluble transcription factor complexes, Molecular Systems Biology 11 (2015) 775.

[104] J. M. Gudas, M. Payton, S. Thukral, E. Chen, M. Bass, M. O. Robinson, S. Coats, Cyclin E2, a novel G1 cyclin that binds Cdk2 and is aberrantly expressed in human cancers, Molecular and Cellular Biology 19 (1999) 612–622.

[105] N. Lauper, A. R. Beck, S. Cariou, L. Richman, K. Hofmann, W. Reith, J. M. Slingerland, B. Amati, Cyclin E2: a novel CDK2 partner in the late G1 and S phases of the mammalian cell cycle, Oncogene 17 (1998) 2637.

[106] M. Zariwala, J. Liu, Y. Xiong, Cyclin E2, a novel human G1 cyclin and activating partner of CDK2 and CDK3, is induced by viral oncoproteins, Oncogene 17 (1998) 2787.

[107] N. Ramachandran, E. Hainsworth, B. Bhullar, S. Eisenstein, B. Rosen, A. Y. Lau, J. C. Walter, J. LaBaer, Self-assembling protein microarrays, Science 305 (2004) 86–90.