UCSF
UC San Francisco Previously Published Works

Title
Reconstructing the Molecular Function of Genetic Variation in Regulatory Networks

Permalink
https://escholarship.org/uc/item/7jz9s2dp

Journal
Genetics, 207(4)

ISSN
0016-6731

Authors
Wilentzik, Roni
Ye, Chun Jimmie
Gat-Viks, Irit

Publication Date
2017-12-01

DOI
10.1534/genetics.117.300381

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Reconstructing the Molecular Function of Genetic Variation in Regulatory Networks

Roni Wilentzik,* Chun Jimmie Ye,† and Irit Gat-Viks*,[1]
*Department of Cell Research and Immunology, Tel Aviv University, Israel 6997801 and †Department of Epidemiology and Biostatistics, University of California, San Francisco, California 94117

ABSTRACT Over the past decade, genetic studies have recognized hundreds of polymorphic DNA loci called response QTLs (reQTLs) as potential contributors to interindividual variation in transcriptional responses to stimulations. Such reQTLs commonly affect the transduction of signals along the regulatory network that controls gene transcription. Identifying the pathways through which reQTLs perturb the underlying network has been a major challenge. Here, we present GEVIN ("Genome-wide Embedding of Variation In Networks"), a methodology that simultaneously identifies a reQTL and the particular pathway in which the reQTL affects downstream signal transduction along the network. Using synthetic data, we show that this algorithm outperforms existing pathway identification and reQTL identification methods. We applied GEVIN to the analysis of murine and human dendritic cells in response to pathogenic components. These analyses revealed significant reQTLs together with their perturbed Toll-like receptor signaling pathways. GEVIN thus offers a powerful framework that renders a comprehensive picture of disease-related DNA loci and their molecular functions within regulatory networks.

KEYWORDS eQTL; gene expression; immune dendritic cells; regulatory networks

REVEALING the genetic basis of interindividual variation and discovering how such variation is manifested in response to environmental changes is a fundamental challenge in molecular genetics. Genome-wide association studies have been highly successful in identifying polymorphic DNA loci associated with physiological or molecular phenotypes (McCarthy *et al.* 2008; Mackay *et al.* 2009). When the phenotype of interest is a quantitative trait, such as blood pressure or cholesterol levels, the underlying genetic locus is referred to as a "QTL". A common strategy investigates the association between quantitative traits of transcriptional responses and their underlying DNA loci called "response QTLs" (reQTLs) (Albert and Kruglyak 2015). Studies have provided clear evidence for the colocalization of reQTLs and disease-related loci (Caliskan *et al.* 2015). This observation has raised the hope that knowledge of reQTLs would contribute to our understanding of genetic variation in regulatory networks that control transcriptional response, and that this, in turn, could shed new light on the molecular basis of disease pathogenesis (Kim and Przytycka 2012; Chasman *et al.* 2016). However, in most reQTL studies the particular positions of genetic perturbations within the regulatory network have not been investigated.

Determining the role of causal reQTLs in the context of regulatory networks is a prerequisite for the experimental investigation of their role in susceptibility to disease. The standard solution to this problem relies on a two-step approach. The first step—"reQTL mapping"—scans the entire genome to identify significant reQTLs, and the second step—"network embedding"—uncovers the position at which an identified reQTL imposes a genetic effect on the network (Tu *et al.* 2006; Suthram *et al.* 2008). Notably, various studies over the last few years have shown that genetic variation can be manifested in the context of a particular cell type or stimulus but not in others (*e.g.*, Petretto *et al.* 2006; Fairfax *et al.* 2014; Kugelberg 2014; De Jager *et al.* 2015). However, the two-step approach has largely ignored the fact that the rewiring of reQTLs in different stimulations is substantial. As a consequence, in its calculation, a fixed (context-independent) reQTL effect on a fixed molecular network is generally assumed.

Copyright © 2017 by the Genetics Society of America
doi: https://doi.org/10.1534/genetics.117.300381
Manuscript received March 14, 2017; accepted for publication September 11, 2017; published Early Online October 10, 2017.
Supplemental material is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.117.300381/-/DC1.
[1]Corresponding author: Rooms 208–209, Department of Cell Research and Immunology, Britannia building, Tel Aviv University, Tel Aviv 6997801, Israel. E-mail: iritgv@post.tau.ac.il

An alternative strategy considers the contextual specificity of both the reQTLs and the regulatory rewiring of signaling pathways. In two recently devised methods—INCIRCUIT and PINE (Gat-Viks *et al.* 2013; Wilentzik and Gat-Viks 2015)—we considered a mechanism in which a causal reQTL affects the transduction of signals along a certain molecular pathway (referred to as a "perturbed network branch"), thereby leading to a perturbed signal propagating along the network, from the extracellular stimulations toward the downstream transcribed genes (Supplemental Material, Figure S1 in File S1). In those methods, the reQTL mapping step exploits standard procedures for the identification of reQTLs, and the network embedding step searches for perturbed network branches by exploiting existing knowledge about the propagation of different stimulation cues through the regulatory network. In particular, PINE was shown to improve on the performance of the INCIRCUIT algorithm, achieving substantially better predictive accuracy (Wilentzik and Gat-Viks 2015). However, analysis using PINE is still limited, most notably owing to its inability to handle heterozygosity and its high false-positive rate. As a result, PINE could not be applied on the outbred human population and is less suitable for accurate identification of reQTLs jointly with their perturbed network branches.

Ideally, a systematic study of the role of reQTLs in regulatory networks should: (i) provide a joint mapping of reQTLs and the position of their effect in networks; (ii) consider the context specificity of both reQTLs and branches within regulatory networks; (iii) achieve high accuracy and reasonable runtime performance; and (iv) handle both homozygous and heterozygous cohorts. To address this fourfold challenge, we have developed a novel method called GEVIN ("Genome-wide Embedding of Variation In Networks") for the simultaneous identification of reQTLs and their perturbed branches within the regulatory network. Notably, GEVIN enables us not only to identify the effect of reQTLs on regulatory networks, but also to provide an *ab initio* identification of reQTLs based on prior knowledge of molecular networks.

To examine the utility of the GEVIN approach, we first evaluated the ability of this method to identify the perturbed network branch. This was done by comparing GEVIN with alternative branch identification methods on simulated data of various network structures. We then compared its performance on synthetic data to those of classical reQTL identification methods, focusing on the ability to correctly identify reQTLs in response to specific stimulations. On all benchmarks, and when using different genetic backgrounds (either heterozygous or homozygous individuals), GEVIN displayed substantial improvements in performance and maintained reasonable running times. Thus, by employing GEVIN it is possible both to utilize information about the regulatory network for enhancement of reQTL mapping and to pinpoint the position of reQTLs within regulatory networks. As a proof of concept, we applied GEVIN on transcriptome data sets to identify the embedding of reQTLs in regulatory networks of the human and of the murine immune systems. Notably, the results suggested possible relationships among viral infections and the pathogenesis of certain inflammatory autoimmune diseases and neurological disorders. Overall, our study highlights the utility of GEVIN in uncovering molecular mechanisms by which polymorphic DNA loci induce broad effects on the interindividual variation in regulatory networks.

Methods

Overview of GEVIN: a method for embedding reQTLs in regulatory networks

The GEVIN algorithm is designed to identify "network-dependent reQTLs" perturbing particular branches within a regulatory network. GEVIN takes as input three types of data (Figure 1A): (i) a regulatory network that was acquired from the scientific literature, consisting of known environmental cues (*e.g.*, stimulations x, y, z) and distinct signaling pathways, termed "network branches" (*e.g.*, branches $b1$–$b10$); (ii) genotyping of all polymorphic DNA loci in each individual [single-nucleotide polymorphisms (SNPs); *e.g.*, SNPs α, β, γ, and individuals 1 to 4]; and (iii) transcriptional response profiles of the network's genes (*e.g.*, genes *a*–*f*) following each of the network's stimulations and across all genotyped individuals. GEVIN is designed on the basis of a fundamental assumption that a perturbation of a SNP in a certain branch will lead to genetic variation in those genes that are regulated by the branch, but only in the context of the specific stimulations that act as triggers of this branch. Based on this assumption, each branch is characterized by an "activation signature," a collection of all combinations of upstream stimulations that trigger the branch and all downstream genes regulated by the branch [*e.g.*, $(c-f)\times(y-z)$ for branch $b6$ in Figure 1B]. When testing the perturbation of a certain branch by a specific SNP, the procedure computes a "GEVIN score," which evaluates the observed effect of the SNP across all combinations of genes and stimuli within the activation signature of the branch. A high GEVIN score for a SNP and its best-scoring branch reflects a promising reQTL together with its inferred perturbed network branch. We term such reQTLs network-dependent reQTLs.

Formally, the GEVIN score is evaluated using a constraint-based multivariate regression model. In this model, the explanatory variable is the SNP, the dependent variable is the multivariate transcriptional responses of all stimulations (across all genes and individuals), and the solution is constrained by the activation signatures of the branches. For example, the activation signature of branch $b6$ in the illustration [$(c-f)\times(y-z)$; Figure 1B], combined with the genotypic data of SNP β (Figure 1A; top right), fully explains the transcriptional response measurements (Figure 1A; bottom right): *AA*- and *aa*-carrying individuals differ in their response of genes *c*–*f* following stimulations y and z. The output is therefore a single hypothesis of the effect of reQTL β on branch $b6$ (Figure 1C).

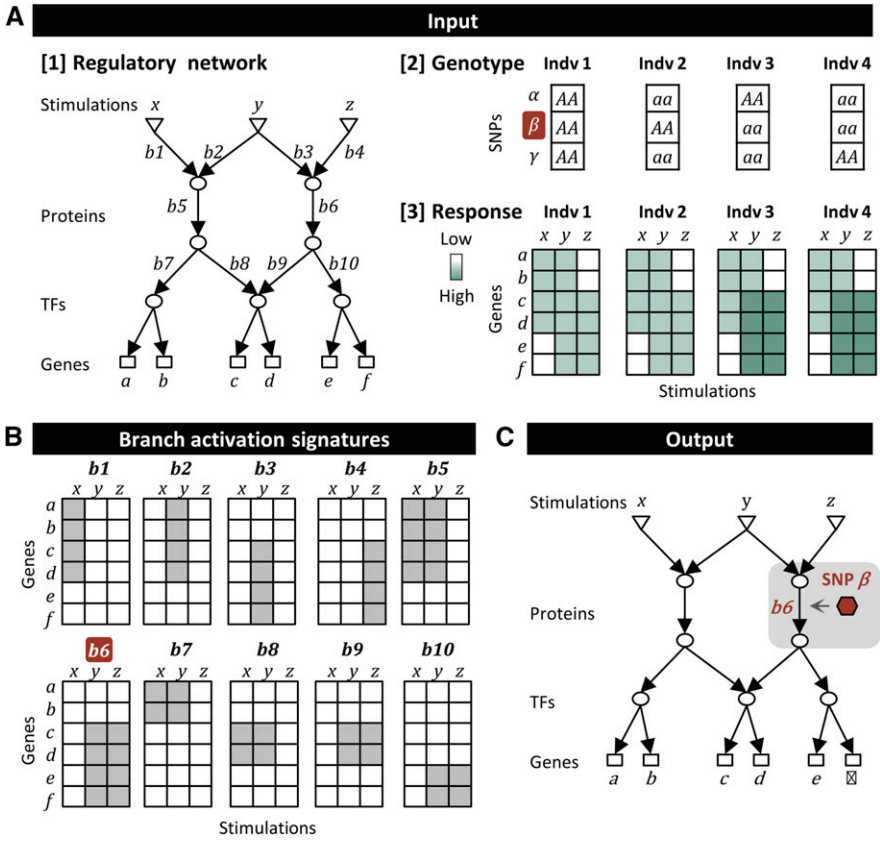


Figure 1 Illustration of the GEVIN algorithm. (A) Input data. GEVIN takes as input [1] a regulatory network consisting of different branches ($b1$–$b10$), [2] genotypic data of multiple DNA polymorphic loci (SNPs α, β, γ) across multiple individuals (Indv, individuals 1 through 4), and [3] transcriptional response levels ("response"; presented on a white−green scale) elicited by the network's triggering stimuli (x, y, z) for all genes regulated by the network (a–f). (B and C) GEVIN's analysis. GEVIN starts by calculating the activation signature of each branch [light gray in (B)], where each activation signature encodes the combination of downstream genes regulated by the branch across the upstream stimulations triggering the signal transduction along the branch [*e.g.*, $(c-f)\times(y-z)$ for branch $b6$]. Next, GEVIN calculates a statistical score for the perturbation of each branch by each SNP. This is done by assessing the agreement between the genetic effect of a SNP [from (A)] with the activation signature of a branch [from (B)]. In particular, genetic variation in SNP β exactly matches the variation in response of genes c–f following stimulations y and z [a higher response in the *aa*-carrying individuals compared to the *AA*-carrying individuals; (A)], which is the activation signature of branch $b6$ [as indicated in (B)]. GEVIN therefore suggests SNP β as a response QTL (reQTL) that leads to transcriptional variation by perturbing the transduction of signals through branch $b6$ [illustrated in (C)]. TF, transcription factor.

Formulation of the GEVIN algorithm

In the following sections, we first formalize the input data, and then describe a standard gene–SNP association model, extend this model for the effect of a single SNP on a single gene through perturbation of a certain network branch, and explain how to model SNP–branch perturbation. Lastly, based on this framework, we describe the details of the GEVIN algorithm as a genome-wide analysis of genetic perturbations within the signaling network.

Input data: GEVIN takes the following as input. (i) Genotyping of a large collection of SNPs Q across a group of individuals I. X^q is a column vector representing the genotyping of SNP $q \in Q$ across all individuals in I. (ii) The transcriptional response of multiple genes in group G following each of the multiple stimulations in group S across all individuals in group I. Let Y^g be the measured $|I| \times |S|$ transcriptional response matrix of gene $g \in G$, where $Y^g_{i,s}$ represents the change in expression level of gene g following stimulation $s \in S$ in individual $i \in I$. (iii) Regulatory network $M^{G,S}$, which is acquired from the scientific literature. Such a network consists of a wiring diagram M that is triggered by the collection of stimulations S and controls the transcriptional response of genes in group G. The structure of the regulatory network M defines a collection of network branches (that is, the distinct signaling pathways), denoted by B. Each branch $b \in B$ is characterized by the subset of its triggering stimulations $S^b \in S$ (its upstream stimuli) and the subset of its regulated genes $G^b \in G$ (its downstream genes). In accordance, we define the activation signature of a branch $b \in B$ as a Boolean $|G| \times |S|$ matrix A^b, where for each $g \in G^b$ and $s \in S^b$, $A^b_{g,s} = 1$ and all other entries are set to zero.

A basic gene–SNP association framework: Our analysis generally relies on a standard multivariate regression model (Rencher 2002; Hidalgo and Goodman 2013) to evaluate the association between a single SNP $q \in Q$ and a single gene $g \in G$ based on its transcriptional response to all stimulations in group S. In this model, the explanatory variable is the genotypic data X^q and the multivariate outcome variable is the transcriptional response matrix Y^g:

$$Y^g_{|I|\times|S|} = \mu + X^q_{|I|\times 1}\beta^{q,g}_{1\times|S|} + \varepsilon \quad \text{(Equation 1)}$$

To gain flexibility in the relationships between the different genotypic groups, the genetic effect was modeled as a categorical variable.

Modeling the effect of a single SNP on a single gene through perturbation of a single network branch: GEVIN relies on a basic assumption that if a given SNP truly perturbs a certain network branch $b \in B$, it would induce genetic variation only in the genes regulated by this branch ($g \in G^b$) and only following the relevant triggering stimulations ($s \in S^b$),

as they compose the activation signature of branch b. To model the effect of SNP q on gene g through perturbation of a given branch $b \in B$, we add constraints to the regression model in Equation 1 as follows: the coefficients $\beta_s^{q,g}$ are fixed to zero for all genes and stimulations in which $A_{g,s}^b = 0$ (that is, $\beta_s^{q,g} = 0$ if and only if $s \in S \backslash S^b$ or $g \in G \backslash G^b$; Figure S2 in File S1 exemplifies the resulting regression in the case of the toy network in Figure 1A). Finally, a likelihood ratio test of the constrained regression model (comparing the model with and without the SNP variable X^q) provides a P-value for the association of a single gene g and a single SNP q acting through a certain branch b. We refer to this P-value as $P^{q,b,g}$. Overall, $|Q| \times |B| \times |G|$ P-values are calculated for each combination of a gene, a SNP, and a network branch.

Modeling SNP–branch perturbation: A P-value for the genetic perturbation of a branch b by a SNP q, denoted $P^{q,b}$, is calculated for each SNP $q \in Q$ and branch $b \in B$ by combining the P-values of all the genes downstream to the branch using Fisher's combined probability test (Fisher 1925). Formally, for each SNP q and branch b, Fisher's test is applied to combine all P-values in the set $\{P^{q,b,g} \big| g \in G^b\}$

Genome-wide analysis of genetic perturbations within the signaling network: To avoid statistical inflation (*e.g.*, due to correlation among target genes), the combined P-values of each branch b are normalized by the "genomic inflation factor" of the branch, as described in Devlin *et al.* (2001). The genomic inflation factor of a given branch b is defined as the ratio of the median of the observed (real data) combined P-values *vs.* the expected median. Here, the expected combined P-values are calculated using 10 permuted data sets that were generated by shuffling the genotypic data labels of all individuals. In all cases, the median is calculated across the set $\{P^{q,b} \big| q \in Q\}$ of combined P-values (that is, across all SNPs for the same branch). We refer to these adjusted combined P-values as the "GEVIN scores."

Finally, a permutation-based false discovery rate (FDR) is determined for each branch to conservatively identify significant SNP–branch perturbations. This is performed by generating R permuted data sets (by randomly reshuffling the genotypic data labels of individuals; here, $R = 10$), calculating GEVIN scores for all SNPs in each permuted data set, and then computing the FDR of a branch b as the fraction of permuted SNPs that obtained higher GEVIN scores than a certain threshold. Importantly, by shuffling only the genotypic data, the activation signature of each branch is retained, maintaining correlations between the regulating genes and the triggering stimulations of each branch.

Comparing GEVIN to existing methods for identifying the perturbed branch

To assess GEVIN's ability to identify the particular network branch in which molecular components are perturbed by a given reQTL, we compared GEVIN to two previously published methods: (i) The INCIRCUIT algorithm (Gat-Viks *et al.* 2013), a qualitative method whose output is a list of branches hypothesized to be perturbed by an input reQTL based on the network structure, the output of which is not accompanied by any statistical assessment; and (ii) the PINE algorithm (Wilentzik and Gat-Viks 2015), a statistical framework that relies on probabilistic graphical modeling (Koller and Friedman 2009) of the regulatory network. PINE has been shown to outperform the INCIRCUIT algorithm (Wilentzik and Gat-Viks 2015) but is limited to homozygous organisms. All reported results were generated using the optimal parameters of each algorithm as described in Wilentzik and Gat-Viks (2015) (PINE: 500 permutations and a confidence level of 0.95; and INCIRCUIT: an association cutoff of 0.1 and enrichment cutoff of 0.9).

All compared methods predict the perturbed branch based on the same prior network and measurements, yet they differ in their basic design. Whereas INCIRCUIT and PINE take as input a single significant reQTL, GEVIN was designed to identify the reQTLs simultaneously with branch identification, reducing the reliance on an input reQTL. In our synthetic data analysis, all compared methods utilize the correct reQTLs as input to obtain a fair comparison.

Comparing GEVIN to whole-genome reQTL identification methods

To assess GEVIN's ability to pinpoint the associated reQTL (out of all SNPs), we implemented three commonly used whole-genome reQTL identification methods. Similarly to GEVIN, all compared methods solve separate regression models for each gene and then utilize Fisher's combined probability test to combine the resulting P-values. The alternative methods mainly differ in how they handle multiple stimulations.

MAXSTIM: The MAXSTIM method utilizes a standard regression model and tests for association following each of the stimulations [as implemented in Fairfax *et al.* (2014) and Lee *et al.* (2014)]. Formally, the association between a SNP q and a gene g is tested in the context of each stimulation s:

$$Y_{|I|\times 1}^{g,s} = \mu + X_{|I|\times 1}^{q} \beta^{q,g,s} + \varepsilon \quad \text{(Equation 2)}$$

As a first step, a P-value is calculated for a certain gene following a given stimulation by testing the significance of the genotypic term (that is, with and without the effect of q). Next, for each stimulation and SNP, a statistical score is generated using Fisher's combined method over all genes. For a given SNP q, an overall score is determined by the best-scoring stimulation.

STIMVAR: The STIMVAR method aims to identify reQTLs by evaluating the SNP–stimulation interaction terms [as in Smith and Kruglyak (2008) and Orozco *et al.* (2012)]. For a given gene g and SNP q, the calculation is based on the following model:

$$Y_{(|I|\times|S|)\times 1}^{g} = \mu + X_{(|I|\times|S|)\times 1}^{q} \beta 1 + Z_{(|I|\times|S|)\times 1} \beta 2 + (X^q \times Z)\beta 3 + \varepsilon \quad \text{(Equation 3)}$$

where $X_{(|I|\times|S|)\times 1}^{q}$ is the concatenation of the X^q vector $|S|$ times and Z is a categorical vector of the particular

stimulation, following which the response measurements were taken. The *P*-value of the association between a SNP and a gene is calculated by testing the significance of the interaction term (that is, with and without the interaction term $X^q \times Z$). For a given SNP q, an overall score is determined by combining the *P*-values of all genes using Fisher's method.

MULTIVARIATE: Similarly to GEVIN, the MULTIVARIATE method was designed based on multivariate regression to draw power from the agreement between multiple stimulations [as suggested in Knott and Haley (2000) and Sukhwinder-Singh *et al.* (2012)]. However, unlike GEVIN, MULTIVARIATE is not a branch-driven method and therefore does not impose constraints on the activation signature of branches. Formally, it uses the linear model in Equation 1 without imposing any constraints on the solution space.

Synthetic data analysis

We generated synthetic data for the purpose of comparing the GEVIN algorithm with branch perturbation and reQTL identification methods. See Supplemental Material, File S1 for details about generating the synthetic data and performance analysis.

Analysis of biological data

Mouse data: We applied GEVIN to a previously published data set (Gat-Viks *et al.* 2013) of bone marrow-derived dendritic cells (DCs) from the BXD collection of recombinant-inbred mice strains (Peirce *et al.* 2004), measuring the expression levels of 422 genes in response to three stimulations: Pam3CSK4 (PAM), lipopolysaccharide (LPS), and polyinosinic-polycytidylic acid (poly I:C), across 43 mice strains. Response levels were defined as the expression levels measured after stimulation normalized by the expression in steady state. Genotypic data of 3796 SNPs was downloaded from WebQTL (Wang *et al.* 2003). The murine Toll-like receptor/RIG-1-like receptor (TLR/RLR) signaling network was constructed based on the scientific literature (Richez *et al.* 2009; Kawai and Akira 2010; Fink and Grandvaux 2013; Kawasaki and Kawai 2014; McNab *et al.* 2015; Pandey *et al.* 2015) and encompasses seven distinct signaling branches (Table S1A in File S1). The analysis of GEVIN was separately conducted on five groups of genes (#1–#5); each gene in these groups was embedded in the network under the relevant transcription factors as described in Gat-Viks *et al.* (2013) (see Table S2A in File S1).

Human data: We applied GEVIN to a recently published data set (Lee *et al.* 2014) of peripheral blood monocyte-derived DCs from 676 healthy individuals, measuring the expression of 414 genes in response to three stimulations: LPS, influenza virus lacking the NS1 gene (denoted FLU), and the interferon-β cytokine (IFN-β). Response levels were calculated as in the mouse data. Genotypic data are available for 531 individuals across 642,850 SNPs (Lee *et al.* 2014). Here, we focused on 266 individuals that were genotyped, and from which samples were taken and exposed to all three stimulations (all profiles exhibit good correlation to other individuals). Out of 642,850 genotyped SNPs, we filtered out 46,523 loci in which one of the genotypic groups has an allele frequency $< 1\%$ in the samples. We modeled the structure of the human TLR/RLR signaling network based on a comprehensive literature survey with a focus on human DCs (Richez *et al.* 2009; Kawai and Akira 2010; Fink and Grandvaux 2013; Kawasaki and Kawai 2014; Lee *et al.* 2014; McNab *et al.* 2015; Pandey *et al.* 2015). Overall, the human network model encompasses nine distinct signaling branches (Table S1B in File S1) and is embedded with 18 different genes (Table S2B in File S1). The multivariate regression model (Equation 1) was extended to control for three confounding factors: gender, age, and ethnicity. In particular, we conducted a preprocessing step for each branch: we started by performing principal component analysis on the confounding factors and then used their representation in the principal component space to control the factors in the multivariate regression of each branch in the network. The residuals of the regressions were later used as dependent variables in the regression model when testing the significance of a SNP–branch perturbation.

Data availability

All data sets used in this work are fully presented in the paper. The code for the GEVIN algorithm is available for download at https://github.com/roniwile/GEVIN or http://csgi.tau.ac.il/gevin/.

Results

Comparison of GEVIN with alternative methods for the challenge of branch identification

To examine GEVIN's ability to correctly identify the perturbed network branch, we compared GEVIN with two alternative branch identification methods: INCIRCUIT and PINE (Gat-Viks *et al.* 2013; Wilentzik and Gat-Viks 2015). For this purpose, we utilized synthetic data collections based on multiple regulatory networks and different data parameters, consisting of case data sets of network-dependent reQTLs (each having a genetic effect on a randomly chosen network branch) as well as control data sets without any genetic effect. In particular, we generated two types of synthetic data collections: those using homozygous and those using heterozygous genetic data (see *Methods*). The INCIRCUIT and PINE methods take as input a known reQTL and focus on using the data to identify the perturbed branch. Accordingly, all three compared algorithms were provided with true synthetic reQTLs as input.

We compared the performances of the three methods on synthetic data sets across a broad range of numbers of genes, numbers of individuals, and genetic effect sizes. Since INCIRCUIT is a qualitative method and does not provide statistical

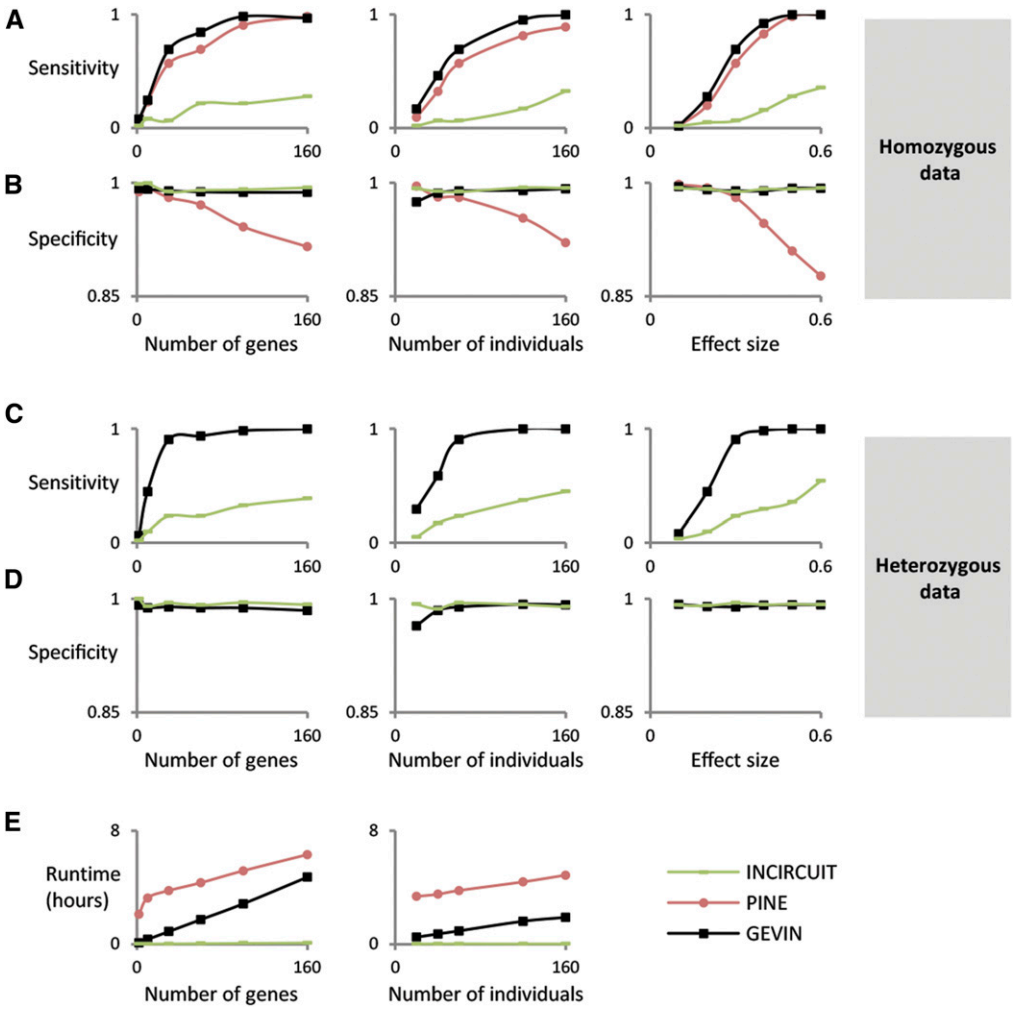


Figure 2 Performance analysis of identifying perturbed network branches. (A–D) Shown are the sensitivity (A and C) and specificity scores (B and D) across different data parameters (numbers of genes, numbers of individuals, and effect sizes; left to right) obtained by three methods for embedding response QTLs (reQTLs) in networks (color coded) using data sets of either homozygous (A and B) or heterozygous (C and D) genetic background. PINE, which cannot be applied on heterozygous data, was excluded from plots C and D. (E) Shown is the runtime (*y*-axis; hours) of the three reQTL embedding algorithms (color coded) across different numbers of genes or individuals using data sets of homozygous genetic background.

scores, we assessed the performance by means of standard sensitivity and specificity statistics, while using a statistical threshold for the PINE and GEVIN methods (see *Methods*). Figure 2A demonstrates the superiority of the GEVIN and PINE methods over INCIRCUIT in correctly identifying the perturbed network branch when tested on homozygous data. Figure 2B highlights a major drawback of the PINE algorithm, which largely fails to identify nonperturbed branches. In particular, PINE is limited in its ability to discriminate between successive network branches, leading to high false-positive identification of branches on the same signaling pathway as the perturbed branch (Figure S3, A and B in File S1). Similar results for PINE and GEVIN were reproduced in various threshold settings using a standard area under the precision-recall curve statistic (denoted "AUPR") across changing score thresholds (Figure S3C in File S1). The superiority of GEVIN is further manifested in data of heterozygous genetic background (Figure 2, C and D): GEVIN outperforms INCIRCUIT, while PINE cannot be applied to heterozygous data.

Encouraged by GEVIN's ability to identify network-dependent reQTLs, we next asked whether GEVIN can further distinguish true network-dependent reQTLs, which affect all genes regulated by the perturbed branch, from single-gene reQTLs, which have an effect on a single gene that is regulated by the perturbed branch. To address this, we compared two types of control data sets: either the original control, in which none of the target genes is associated with upstream reQTLs, or an alternative control data set, in which only a single downstream gene is associated with a randomly chosen reQTL (*Methods*). In both cases, we used the original network-dependent reQTLs as our case data sets. As demonstrated in Figure S4 in File S1, GEVIN's ability to distinguish network-dependent reQTLs from single-gene reQTLs is similar to its ability to distinguish network-dependent reQTLs from non-reQTLs.

Finally, we examined the runtime performances of the three branch identification methods. We did this by comparing the homozygous data sets across varying numbers of genes and individuals on a Linux machine with 2.6 GHz AMD Opteron 6238 processors. As shown in Figure 2E, GEVIN maintains a reasonable running time that is lower than PINE's. For example, PINE's runtime is four times higher than GEVIN's when using 30 genes and 2.5 times higher when using 60 genes.

Taken together, these results demonstrate the advantages of GEVIN as a branch identification method: it is more accurate than INCIRCUIT and PINE, and can be applied to a wide variety of data sets of both homozygous and heterozygous

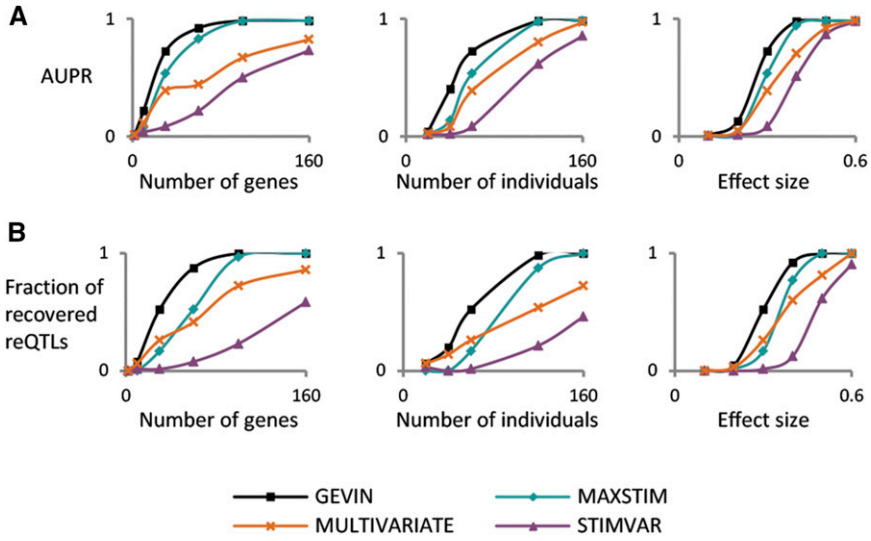


Figure 3 Comparative analysis of response QTL (reQTL) identification methods. Shown are the precision-recall curve statistic (AUPR) scores [*y*-axis; (A)] and the fraction of recovered reQTLs [*y*-axis; (B)] using the case data sets of network-dependent reQTLs that were constructed with different numbers of genes (left), different numbers of individuals (middle), and different effect sizes (right; *x*-axis; 100 heterozygous SNPs in all cases). The four reQTL identification methods are color coded. The plots depict GEVIN as the best performing method for identifying network-dependent reQTLs.

genetic backgrounds. In the following section, we describe further testing of GEVIN's performance as a reQTL identification methodology.

Comparison of GEVIN with alternative methods for the challenge of reQTL identification

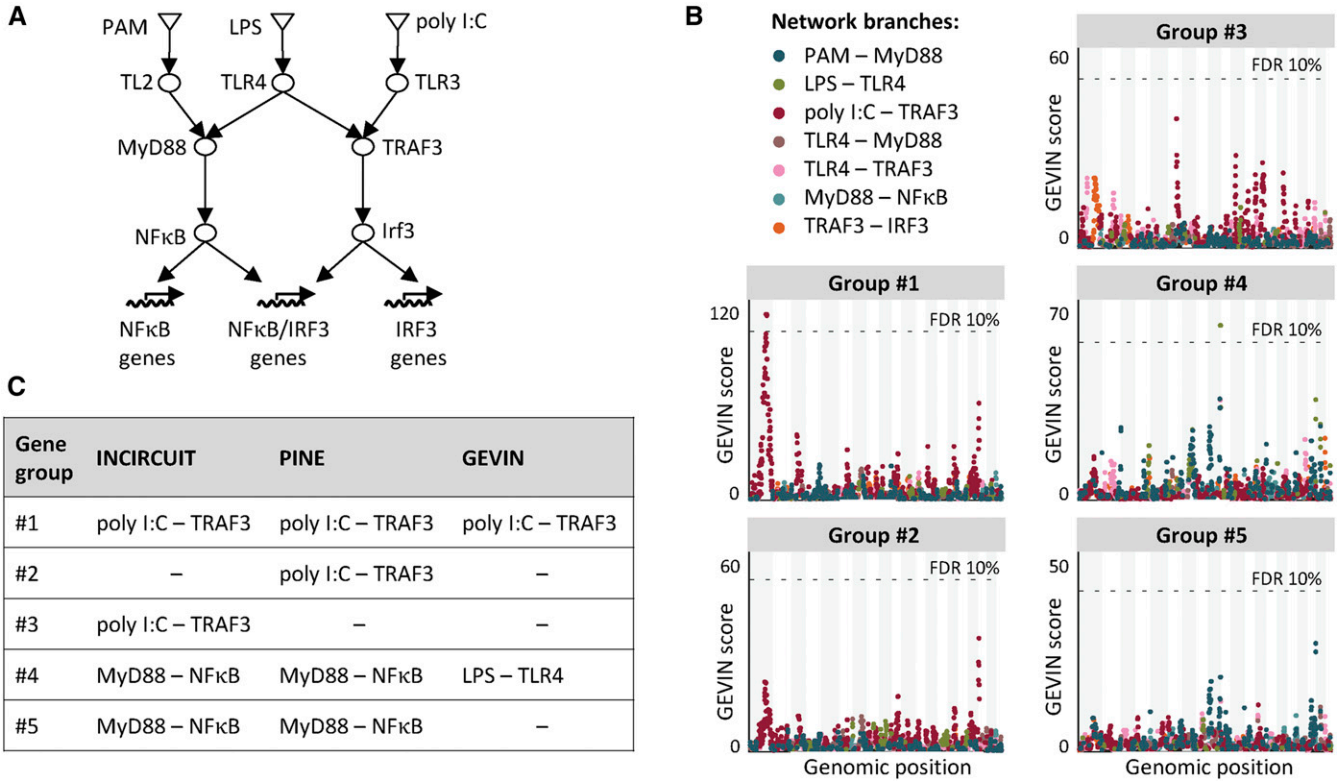
To characterize the ability of GEVIN to identify the true reQTL among all candidate SNPs, we compared GEVIN with three alternative regression-based methods—MAXSTIM, STIMVAR, and MULTIVARIATE—that differ from each other and from GEVIN in handling the transcriptional responses to multiple stimulations (see *Methods*). MAXSTIM is based on regression analysis performed separately for each of the stimulations (Fairfax *et al.* 2014; Lee *et al.* 2014); STIMVAR evaluates the significance of interactions between a given SNP and the various stimulations (Smith and Kruglyak 2008; Orozco *et al.* 2012); and MULTIVARIATE utilizes multivariate regression to exploit correlations between multiple stimulations (Knott and Haley 2000; Sukhwinder-Singh *et al.* 2012). Note that both GEVIN and MULTIVARIATE handle multiple stimulations as a multivariate outcome, but they use different sets of genes and stimuli; whereas MULTIVARIATE considers all genes and stimulations in the network, GEVIN is focused on the subgroup of genes and stimulations relevant to a particular branch within the network.

We compared the performances of these methods by using the abovementioned case data sets of network-dependent reQTLs, where each data set was built on the basis of one randomly chosen reQTL from a synthetic genome of 100 SNPs. Data sets without any genetic effect were used as control. Given the imbalance between the numbers of synthetic causal reQTLs and of the remaining SNPs, we tested the performance of the reQTL identification methods by using the AUPR statistic. We further examined the fraction of case data sets in which the synthetic causal reQTLs exactly matched the reQTLs predicted by each of these methods (denoted "fraction of recovered reQTLs"). This analysis demonstrated the advantage of GEVIN in the case of network-dependent reQTLs, where all reQTLs in the data set have an effect on a certain branch within the network (Figure 3). For example, in a synthetic data collection with 30 genes, 60 individuals, and a network-dependent effect size of 0.4, GEVIN's AUPR score was higher than the scores obtained by MAXTIM, MULTIVARIATE, or STIMVAR (0.98 *vs.* 0.94, 0.7, or 0.51, respectively). Similar results were obtained for the fraction of recovered reQTL scores across different numbers of genes, numbers of individuals, and genetic effect sizes.

We were also interested in examining the performance of GEVIN compared with those of alternative reQTL identification methods in the case of network-independent (constitutive) reQTLs. Unlike the effects of network-dependent reQTLs, which act in the context of a subset of the stimulations (depending on the localization of the perturbed branch), the effect of constitutive reQTLs is manifested in all stimulations. Accordingly, we constructed a collection of synthetic case data sets carrying such reQTLs (see *Methods*). Notably, in the case of constitutive reQTLs, the MULTIVARIATE method outperformed the alternative methods, including GEVIN (Figure S5 in File S1). This finding is largely consistent with the designs of the GEVIN and MULTIVARIATE methods, which exploit the commonality between various stimulations. However, whereas MULTIVARIATE exploits all stimulations as a multivariate outcome, GEVIN typically tests a subset of the stimulations, an appropriate constraint only in the case of network-dependent reQTLs. A demonstration of the differences between the four methods in the case of both network-dependent and constitutive reQTLs is further provided in Figure S6 in File S1. Taken together, these results show that GEVIN successfully utilizes the differences in genetic associations between stimulations, performing best in the case of network-dependent reQTLs.

Application of GEVIN to the examination of murine and human immune DCs

To study network-dependent reQTLs in real biological data, we used GEVIN to examine reQTLs related to the response of both human and murine primary DCs to pathogenic



C

Gene group	INCIRCUIT	PINE	GEVIN
#1	poly I:C – TRAF3	poly I:C – TRAF3	poly I:C – TRAF3
#2	–	poly I:C – TRAF3	–
#3	poly I:C – TRAF3	–	–
#4	MyD88 – NFκB	MyD88 – NFκB	LPS – TLR4
#5	MyD88 – NFκB	MyD88 – NFκB	–

Figure 4 Evaluation of murine dendritic cell (DC) data using GEVIN. (A) Model of the murine Toll-like receptor/RIG-1-like receptor (TLR/RLR) signaling network. The network is triggered by three extracellular stimulations [Pam3CSK4 (PAM), lipopolysaccharide (LPS), and polyinosinic-polycytidylic acid (poly I:C)] and regulates the transcription of NFκB-induced or IRF3-induced genes or both (see gene embedding in Table S2A in File S1). (B) Application of GEVIN on the five groups of genes (groups #1−#5). Presented are the GEVIN scores (*y*-axis) obtained for each SNP (*x*-axis; gray background for odd chromosomes) for the best-scoring perturbed branch (color coded). Horizontal line: the most conservative 10% permutation-based false discovery rate (FDR) threshold across all branches. (C) Comparison of GEVIN with alternative network embedding methods. Shown for each gene group (column 1) are the branches predicted by the INCIRCUIT (column 2), PINE (column 3) and GEVIN (FDR < 10%; column 4) algorithms.

components. In each case, we first discuss the data set and regulatory network and then consider the resulting associations.

We first applied GEVIN to the examination of bone marrow-derived immune DCs of recombinant inbred BXD mouse strains (Peirce *et al.* 2004). We used previously published data (Gat-Viks *et al.* 2013) on transcriptional responses to three pathogenic-like *ex vivo* stimulations: PAM, LPS, and poly I:C. Each of these stimulations triggers the TLR/RLR signaling network, which plays a key role in pathogen recognition of the innate immune system (Richez *et al.* 2009; Kawai and Akira 2010; Fink and Grandvaux 2013; Kawasaki and Kawai 2014; McNab *et al.* 2015; Pandey *et al.* 2015). To apply GEVIN to these data we modeled the murine TLR/RLR signaling network (depicted in Figure 4A and described in *Methods*). We applied GEVIN to five groups of genes (groups #1 to #5) that were previously characterized in immune DCs (Gat-Viks *et al.* 2013; Wilentzik and Gat-Viks 2015) and were scrutinized for the embedding of their associated reQTLs in the TLR/RLR signaling network. According to GEVIN's predictions (Figure 4B and Figure S7 in File S1), two reQTLs (located in chromosome 1: 128−185 Mb and chromosome 9: 122−123 Mb) perturbed two distinct branches in the TLR/RLR signaling network (poly I:C–TRAF3 and LPS–TLR4, respectively), leading to genetic variation in the response of genes in groups #1 and #4, respectively (permutation-based FDR < 10%). Overall, the predicted reQTLs of all groups were in agreement with previous studies (Figure 4C and Figure S8 in File S1). Specifically, the branch predicted by GEVIN to perturb the genes in group #1 is consistent with the branches suggested by INCIRCUIT and PINE, while the branch predicted by GEVIN for group #4 is upstream of the branch suggested by the former methods, sharing part of its signal transduction pathways. The best-scoring branches suggested by GEVIN for all other groups (#2, #3, and #5) largely agreed with the branches identified by INCIRCUIT and PINE, although those branches did not pass the 10% FDR threshold.

We next applied GEVIN to the recently published human DC data set (Lee *et al.* 2014), which records transcriptional responses to three *ex vivo* stimulations (IFN-β, LPS, and FLU) that trigger the TLR/RLR signaling network. The regulatory network used for this analysis is illustrated in Figure 5A (*Methods*). A summary of GEVIN's predictions across all branches is shown in Figure 5B (see individual branches in Figure S9 in File S1), depicting a significant reQTL that perturbs the FLU–IFN-α branch (SNP rs327028, located in chromosome 11: 20792419 bp, permutation-based FDR < 10%). According to the activation signature of the FLU–IFN-α

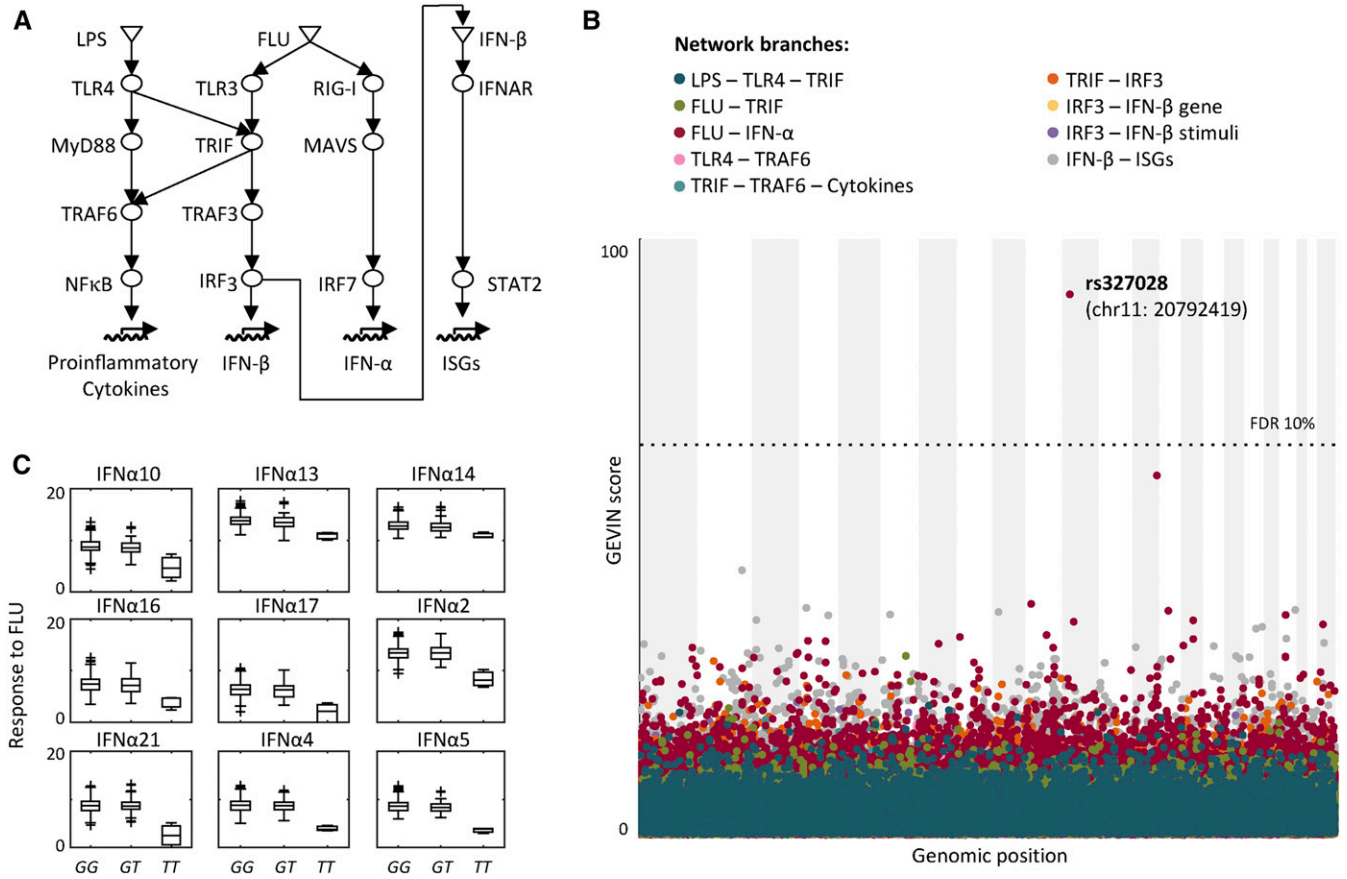


Figure 5 Evaluation of human dendritic cell (DC) data using GEVIN. (A) Model of the human Toll-like receptor/RIG-1-like receptor (TLR/RLR) signaling network. The network is triggered by three stimulations [lipopolysaccharide (LPS), influenza virus lacking the NS1 gene (FLU), and interferon (IFN)-β] and regulates the transcription of downstream genes [proinflammatory cytokines, IFN-α and -β, and IFN-stimulated genes (ISGs); see gene embedding in Table S2B in File S1]. (B) Presented are the GEVIN scores (*y*-axis) obtained for each SNP (*x*-axis; gray background for odd chromosomes) for the best-scoring perturbed branch (color coded). Horizontal line: the most conservative 10% permutation-based false discovery rate (FDR) threshold across all branches. The analysis reveals a single significant perturbation of response QTL (reQTL) rs327028 (located on chromosome 11: 20792419 bp), which likely perturbs the signal transduction from the FLU stimulation to the IFN-α genes. (C) Shown are boxplots of the transcriptional FLU-responses (*y*-axis) of the nine IFN-α genes. The response levels are depicted across the three genotypic groups (*x*-axis) derived from the rs327028 SNP. The plots indicate that transcriptional responses, particularly of IFN-α genes, are lowest in *TT*-carrying individuals.

branch, rs327028 should have an influenza-specific effect on IFN-α genes (the effect would not be observed in response to IFN-β or LPS). This agrees well both with the results of a simple ANOVA model, where the SNP specifically affected the suggested target genes after influenza stimulation (Figure S10 in File S1), and with the lower expression levels observed in the minor allele homozygous group (*TT*; Figure 5C). The suggested rs327028 reQTL has a minor allele frequency of 10.14% (dbSNP; Sherry *et al.* 2001), does not deviate significantly from Hardy–Weinberg equilibrium, and is located in the first exon of the short isoform of the *NELL1* gene (Pang *et al.* 2015). We conclude that this SNP constitutes a previously uncharacterized reQTL that affects the transcriptional response of IFN-α genes in immune DCs following influenza infection.

Discussion

Extensive efforts over the past decade have been focused on uncovering the genetic basis of complex phenotypes. Nevertheless, the molecular functions of reQTLs that are involved in many diseases are still largely unknown. Recent studies have focused either on the identification of reQTLs (reQTL mapping; Lee *et al.* 2014; Caliskan *et al.* 2015) or on identification of the position of a given reQTL within the molecular network (network embedding; Tu *et al.* 2006; Suthram *et al.* 2008; Gat-Viks *et al.* 2013; Wilentzik and Gat-Viks 2015). The GEVIN algorithm is the first to combine these two tasks by simultaneously identifying reQTLs and their positions within the underlying regulatory network. According to our results with synthetic data, when a reQTL has a specific effect on a certain pathway within a regulatory network (network-dependent reQTL), GEVIN outperforms existing methods both of network embedding and of reQTL mapping (Figure 2 and Figure 3). Our application of GEVIN on a well-characterized data set of murine immune response demonstrated that GEVIN's predictions are consistent with previously published results (Figure 4), supporting the validity of the GEVIN algorithm.

Finally, we applied GEVIN on a recently published data set of human immune responses to LPS, IFN-β, and influenza stimulations (Figure 5). A particularly promising finding concerned the human rs327028 SNP, which is located on chromosome 11 within the *NELL1* gene and is associated with transcriptional responses of several IFN-α genes. The effect of this SNP was specific to influenza stimulation, and was not observed in either LPS or IFN-β stimulations. The mechanism proposed by the GEVIN algorithm therefore implies that *NELL1* participates in the regulation of IFN-α levels during the cellular response to viral pathogens. In support of this hypothesis, previous studies have demonstrated the involvement of *NELL1* in several immune-related disorders. In particular, a polymorphic site within the *NELL1* gene was reported to have a genetic effect on Crohn's disease (Franke *et al.* 2007), one of the two subphenotypes of inflammatory bowel disease (IBD). IBD is an autoimmune disease triggered by interactions of the intestine with environmental factors such as viral infections. In another study, examination of changes in IFN-α levels in response to the Newcastle disease virus revealed that the kinetics of IFN-α production in Crohn's patients differs from that in controls (Capobianchi *et al.* 1992). Altogether, those studies showed that *NELL1* is related to IBD and that virus-infected cells from IBD patients exhibit an altered IFN-α response. This observation can be explained by GEVIN's proposed mechanism of *NELL1* regulating IFN-α genes as a protective response against viral attack. Further study is needed to understand whether IFN-α indeed mediates between *NELL1* activity and IBD symptoms in response to viral infection.

Another interesting finding regarding the involvement of *NELL1* in IFN-α regulation appears in the literature on neurological disorders. *NELL1* resides in a genomic locus previously associated with a behavioral phenotype in autism spectrum disorders (ASDs) of faints, fits, or blackouts (Connolly *et al.* 2013), and with fever-related syndromic epilepsy in ASD patients (Hartmann *et al.* 2015). Several items of evidence have provided intriguing indications for the role of IFN-α in ASD, including raised IFN-α levels in patients with ASD (Stubbs 1995) and autistic-like symptoms (*e.g.*, "withdrawal" and "reduced communication abilities") following a large dose of IFN-α given to children as a treatment for cancer (Hill *et al.* 1981). Interestingly, it was also suggested that high IFN-α levels are related to seizures, which are prevalent in ASD patients (Parmeggiani *et al.* 2010). Finally, a recent study identified the TLR signaling network as one of the two signaling pathways most significantly shared between ASD and its comorbidities (Nazeen *et al.* 2016). Altogether, our research raises the possibility that ASD—and particularly seizure pathogenesis—might be affected by the control of *NELL1* on IFN-α levels in response to viral infections.

The GEVIN algorithm was designed to exploit potential cross talk among different overlapping signaling pathways in response to different stimulations. This relies on a fundamental assumption that most biological pathways are a part of a more complex molecular network and that overlapping pathways are common in these networks. We base this assumption on the high prevalence of overlapping pathways in the various repositories of manually-curated signaling networks. For instance, overlapping pathways are present in 65% (15 of 23) of the multi-stimuli cancer networks and 90% (10 of 11) of the multi-stimuli cardiovascular networks that were annotated in the Ingenuity Pathway Analysis database (QIAGEN, Valencia, CA). We hope that future experiments will facilitate the construction of additional, less explored networks.

The GEVIN approach opens up many directions for future improvements. First, integration of different types of networks—such as metabolic networks and cell–cell interaction networks—may confer a different perspective on gene regulation. Second, it is likely that the GEVIN framework can be extended to handle other types of reQTLs, such as DNA methylation QTL. Third, integration of nonregulated genes in the computation of branches might allow us to better distinguish between consecutive branches along the same signaling pathway. Finally, simultaneous investigation of several reQTLs affecting the same regulatory network might allow us to explain the appearance of genetic interactions in some target genes (residing downstream of both reQTLs) but not in other genes (residing downstream of only one reQTL). Overall, GEVIN is suggested as a promising approach for investigating reQTLs and modeling their effects on regulatory networks, thus providing an opportunity to obtain a comprehensive view of the molecular mechanisms involved in the pathogenesis of complex diseases.

Acknowledgments

This research was supported by Israel Science Foundation funding (grant number 1643/13), and by fellowships from the Colton Family Foundation, the Edmond J. Safra Center for Bioinformatics at Tel-Aviv University, and the Israeli Centers of Research Excellence (I-CORE) Gene Regulation in Complex Human Disease, Center number 41/11 (R.W.). I.G.-V. is a Faculty Fellow of the Edmond J. Safra Center for Bioinformatics at Tel Aviv University and an Alon Fellow.

Literature Cited

Albert, F. W., and L. Kruglyak, 2015 The role of regulatory variation in complex traits and disease. Nat. Rev. Genet. 16: 197–212.

Caliskan, M., S. W. Baker, Y. Gilad, and C. Ober, 2015 Host genetic variation influences gene expression response to rhinovirus infection. PLoS Genet. 11: e1005111.

Capobianchi, M. R., S. Fais, F. Mercuri, M. Boirivant, F. Dianzani *et al.*, 1992 Interferon-alpha (IFN-alpha) production by human intestinal mononuclear cells. Response to virus in control subjects and in Crohn's disease. Gut 33: 897–901.

Chasman, D., A. Fotuhi Siahpirani, and S. Roy, 2016 Network-based approaches for analysis of complex biological systems. Curr. Opin. Biotechnol. 39: 157–166.

Connolly, J. J., J. T. Glessner, and H. Hakonarson, 2013 A genome-wide association study of autism incorporating autism diagnostic interview-revised, autism diagnostic observation schedule, and social responsiveness scale. Child Dev. 84: 17–33.

De Jager, P. L., N. Hacohen, D. Mathis, A. Regev, B. E. Stranger *et al.*, 2015 ImmVar project: insights and design considerations for future studies of "healthy" immune variation. Semin. Immunol. 27: 51–57.

Devlin, B., K. Roeder, and L. Wasserman, 2001 Genomic control, a new approach to genetic-based association studies. Theor. Popul. Biol. 60: 155–166.

Fairfax, B. P., P. Humburg, S. Makino, V. Naranbhai, D. Wong *et al.*, 2014 Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science 343: 1246949.

Fink, K., and N. Grandvaux, 2013 STAT2 and IRF9: beyond ISGF3. JAK-STAT 2: e27521.

Fisher, R. A., 1925 *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.

Franke, A., J. Hampe, P. Rosenstiel, C. Becker, F. Wagner *et al.*, 2007 Systematic association mapping identifies NELL1 as a novel IBD disease gene. PLoS One 2: e691.

Gat-Viks, I., N. Chevrier, R. Wilentzik, T. Eisenhaure, R. Raychowdhury *et al.*, 2013 Deciphering molecular circuits from genetic variation underlying transcriptional responsiveness to stimuli. Nat. Biotechnol. 31: 342–349.

Hartmann, C., S. von Spiczak, A. Suls, S. Weckhuysen, G. Buyse *et al.*, 2015 Investigating the genetic basis of fever-associated syndromic epilepsies using copy number variation analysis. Epilepsia 56: e26–e32.

Hidalgo, B., and M. Goodman, 2013 Multivariate or multivariable regression? Am. J. Public Health 103: 39–40.

Hill, N. O., A. Pardue, A. Kahn, D. Aleman, and J. M. Hill, 1981 High-dose human leukocyte interferon trials in leukemia and cancer. Med. Pediatr. Oncol. 9: 82.

Kawai, T., and S. Akira, 2010 The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. Nat. Immunol. 11: 373–384.

Kawasaki, T., and T. Kawai, 2014 Toll-like receptor signaling pathways. Front. Immunol. 5: 461.

Kim, Y. A., and T. M. Przytycka, 2012 Bridging the gap between genotype and phenotype via network approaches. Front. Genet. 3: 227.

Knott, S. A., and C. S. Haley, 2000 Multitrait least squares for quantitative trait loci detection. Genetics 156: 899–911.

Koller, D., and N. Friedman, 2009 *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, Cambridge, MA.

Kugelberg, E., 2014 Immunogenetics: tracking immune activity across the genome. Nat. Rev. Immunol. 14: 212.

Lee, M. N., C. Ye, A. C. Villani, T. Raj, W. Li *et al.*, 2014 Common genetic variants modulate pathogen-sensing responses in human dendritic cells. Science 343: 1246980.

Mackay, T. F., E. A. Stone, and J. F. Ayroles, 2009 The genetics of quantitative traits: challenges and prospects. Nat. Rev. Genet. 10: 565–577.

McCarthy, M. I., G. R. Abecasis, L. R. Cardon, D. B. Goldstein, J. Little *et al.*, 2008 Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat. Rev. Genet. 9: 356–369.

McNab, F., K. Mayer-Barber, A. Sher, A. Wack, and A. O'Garra, 2015 Type I interferons in infectious disease. Nat. Rev. Immunol. 15: 87–103.

Nazeen, S., N. P. Palmer, B. Berger, and I. S. Kohane, 2016 Integrative analysis of genetic data sets reveals a shared innate immune component in autism spectrum disorder and its co-morbidities. Genome Biol. 17: 228.

Orozco, L. D., B. J. Bennett, C. R. Farber, A. Ghazalpour, C. Pan *et al.*, 2012 Unraveling inflammatory responses using systems genetics and gene-environment interactions in macrophages. Cell 151: 658–670.

Pandey, S., T. Kawai, and S. Akira, 2015 Microbial sensing by Toll-like receptors and intracellular nucleic acid sensors. Cold Spring Harb. Perspect. Biol. 7: a016246.

Pang, S., J. Shen, Y. Liu, F. Chen, Z. Zheng *et al.*, 2015 Proliferation and osteogenic differentiation of mesenchymal stem cells induced by a short isoform of NELL-1. Stem Cells 33: 904–915.

Parmeggiani, A., G. Barcia, A. Posar, E. Raimondi, M. Santucci *et al.*, 2010 Epilepsy and EEG paroxysmal abnormalities in autism spectrum disorders. Brain Dev. 32: 783–789.

Peirce, J. L., L. Lu, J. Gu, L. M. Silver, and R. W. Williams, 2004 A new set of BXD recombinant inbred lines from advanced intercross populations in mice. BMC Genet. 5: 7.

Petretto, E., J. Mangion, N. J. Dickens, S. A. Cook, M. K. Kumaran *et al.*, 2006 Heritability and tissue specificity of expression quantitative trait loci. PLoS Genet. 2: e172.

Rencher, A. C., 2002 *Methods of Multivariate Analysis* Wiley-Interscience, New York.

Richez, C., K. Yasuda, A. A. Watkins, S. Akira, R. Lafyatis *et al.*, 2009 TLR4 ligands induce IFN-alpha production by mouse conventional dendritic cells and human monocytes after IFN-beta priming. J. Immunol. 182: 820–828.

Sherry, S. T., M. H. Ward, M. Kholodov, J. Baker, L. Phan *et al.*, 2001 dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29: 308–311.

Smith, E. N., and L. Kruglyak, 2008 Gene-environment interaction in yeast gene expression. PLoS Biol. 6: e83.

Stubbs, G., 1995 Interferonemia and autism. J. Autism Dev. Disord. 25: 71–73.

Sukhwinder-Singh, M. V., J. Hernandez, P. K. Crossa, N. S. Singh, Bains *et al.*, 2012 Multi-trait and multi-environment QTL analyses for resistance to wheat diseases. PLoS One 7: e38008.

Suthram, S., A. Beyer, R. M. Karp, Y. Eldar, and T. Ideker, 2008 eQED: an efficient method for interpreting eQTL associations using protein networks. Mol. Syst. Biol. 4: 162.

Tu, Z., L. Wang, M. N. Arbeitman, T. Chen, and F. Sun, 2006 An integrative approach for causal gene identification and gene regulatory pathway inference. Bioinformatics 22: e489–e496.

Wang, J., R. W. Williams, and K. F. Manly, 2003 WebQTL: web-based complex trait analysis. Neuroinformatics 1: 299–308.

Wilentzik, R., and I. Gat-Viks, 2015 A statistical framework for revealing signaling pathways perturbed by DNA variants. Nucleic Acids Res. 43: e74.

Communicating editor: C. Kendziorski