

RESEARCH ARTICLE

# Discovery of Intermediary Genes between Pathways Using Sparse Regression

Kuo-ching Liang, Ashwini Patil, Kenta Nakai\*

Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan

\* [knakai@ims.u-tokyo.ac.jp](mailto:knakai@ims.u-tokyo.ac.jp)



OPEN ACCESS

**Citation:** Liang K-c, Patil A, Nakai K (2015) Discovery of Intermediary Genes between Pathways Using Sparse Regression. PLoS ONE 10(9): e0137222. doi:10.1371/journal.pone.0137222

**Editor:** Niranjana Baisakh, Louisiana State University Agricultural Center, UNITED STATES

**Received:** February 23, 2015

**Accepted:** August 14, 2015

**Published:** September 8, 2015

**Copyright:** © 2015 Liang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The time-course mouse dendritic cell RNA-Seq data is available in the Sequence Read Archive under the accession number DRA001131. The HapMap Caucasian individual RAN-Seq data is available at the ReCount website: <http://bowtie-bio.sourceforge.net/recount/>.

**Funding:** This research is funded in part by Research on Regenerative Medicine for Clinical Application, Health and Labour Sciences Research Grants (<https://research-er.jp/projects/view/885791>). This research is also funded by the Cabinet Office, Government of Japan and the Japan Society for the Promotion of Science (JSPS) through the Funding Program for World-Leading Innovation R&D on

## Abstract

The use of pathways and gene interaction networks for the analysis of differential expression experiments has allowed us to highlight the differences in gene expression profiles between samples in a systems biology perspective. The usefulness and accuracy of pathway analysis critically depend on our understanding of how genes interact with one another. That knowledge is continuously improving due to advances in next generation sequencing technologies and in computational methods. While most approaches treat each of them as independent entities, pathways actually coordinate to perform essential functions in a cell. In this work, we propose a methodology based on a sparse regression approach to find genes that act as intermediary to and interact with two pathways. We model each gene in a pathway using a set of predictor genes, and a connection is formed between the pathway gene and a predictor gene if the sparse regression coefficient corresponding to the predictor gene is non-zero. A predictor gene is a shared neighbor gene of two pathways if it is connected to at least one gene in each pathway. We compare the sparse regression approach to Weighted Correlation Network Analysis and a correlation distance based approach using time-course RNA-Seq data for dendritic cell from wild type, MyD88-knockout, and TRIF-knockout mice, and a set of RNA-Seq data from 60 Caucasian individuals. For the sparse regression approach, we found overrepresented functions for shared neighbor genes between TLR-signaling pathway and antigen processing and presentation, apoptosis, and Jak-Stat pathways that are supported by prior research, and compares favorably to Weighted Correlation Network Analysis in cases where the gene association signals are weak.

## Introduction

Genes in eukaryotic genomes rarely work alone, rather, they cooperate and interact with other genes to form networks or pathways. Gene products can act as activators or repressors to other genes, or bind with each other to form more complicated structures. Many types of pathways or interaction databases have been made available, such as databases for metabolic pathways [1, 2], signal transduction pathways [3, 4], and protein-protein interaction networks [5, 6].

Science and Technology (FIRST Program) for the Comprehensive understanding of immune dynamism: toward manipulation of immune responses (<https://research-er.jp/projects/view/106146>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

The use of pathways in gene, protein, and genome structural variation analyses has become increasingly important as our understanding of the networks and pathways improved with recent advances in high-throughput technology. It has allowed researchers to make sense of observations about the expressions of genes or proteins not only as singular events, but also in a broader context of what is occurring in their interaction neighborhoods. Our current approaches for analyzing gene expression profiles using pathways often rely on finding pathways with an overrepresentation of differentially expressed genes [7]. In these types of analyses, a pathway is seen as a collection of genes independent from other genes and pathways. However, pathways often work together as a cascade of pathways for the transduction of biological signals and for other cellular functions. Therefore, while much progress has been made in the understanding of individual pathways, pathway-based analyses are often affected by the interaction or crosstalk that exist between different pathways [8]. Given that current knowledge of gene interactions is still incomplete, many interactions may still exist between upstream and downstream pathways. Finding these genes to fill in the missing pieces of the puzzle is crucial to the full understanding of the interactive pathways in our genome.

Correlation of gene expression is a common approach for finding novel gene interactions [9–11], but it can be sensitive to sample size [12]. In this work, we propose a sparse regression-based methodology, aimed at discovering intermediary genes between two pathways. For the analysis of pathways 1 and 2, the proposed method divides genes into three gene sets: genes in pathway 1, genes in pathway 2, and the remaining genes. It looks for genes in the remaining set that are associated with genes in both pathways, i.e., shared neighbor genes of the two pathways. More specifically, we use sparse regression with the remaining genes as predictors to model genes in pathways 1 and 2. Predictor genes having non-negative coefficients are considered as having interactions with the modeled pathway genes.

For a knockout experiment, we can further compare the shared neighbor genes found in the wild type and knockout samples, in order to find shared neighbor genes that are uniquely affected by the knockouts. Comparison of changes in shared neighbor genes can lead to discovery of genes that are essential for the communication between the two pathways.

## Materials and Methods

In this work, we will formulate the proposed methodology for RNA-Seq gene expression profile experiments. The method is applied to RNA-Seq data to discover shared neighbor genes, but can also be used with any technology that measures the gene expression profile of a sample.

### Dataset

We use two RNA-Seq datasets to evaluate the proposed method for the prediction of shared neighbor genes between two pathways. The first dataset is a time-series gene-knockout experiment of mouse dendritic cells, which was previously made public in [13]. The second set consists of RNA-Seq data of 60 Caucasian individuals [14] obtained from [15].

**Mouse Dendritic Cell Knock-out Time-course Data.** In adaptive immune response, dendritic cells act as intermediary between antigens and mammalian immune mechanism by processing and presenting antigens to lymphocytes. One of the most important pathways involved in the activation of innate immune response is the Toll-like receptor 4 (TLR4) signaling pathway. TLR-4 signaling pathway is activated when lipopolysaccharide (LPS) found on the surface of Gram-negative bacteria is bound to the extracellular domain of TLR4, which eventually leads to the activation of proinflammatory cytokines and type-1 interferons [16]. After LPS binding, TLR4 signaling branches into two pathways, independently utilizing the adaptor proteins MyD88 and TRIF [17]. MyD88-dependent pathway is utilized for the rapid activation of

IRAK1, IRAK4, and TAK1, which are important for the activation of MAPK and NF- $\kappa$ B genes, whereas the TRIF-dependent pathway is essential for the production of interferon- $\beta$  and late-phase activation of NF- $\kappa$ B [18]. Understanding how the two independent pathways interact with downstream activities, and finding genes that are involved in signal transduction between the upstream and downstream pathways, are important steps for further understanding of mammalian adaptive immune response. In this work we use a dataset that consists of wild type, MyD88 KO, and TRIF KO mouse dendritic cell samples. Each sample was extracted from bone-marrow cells under the presence of GM-CSF. All three types of cells were then stimulated with LPS to elicit immune response. Samples from the stimulated cells were collected at 0hr, 0.5hr, 1hr, 2hrs, 3hrs, 4hrs, 6hrs, 8hrs, 16hrs, and 24hrs after stimulation, and RNA-Seq was performed on each sample. The time-series RNA-Seq data is currently available in Sequence Read Archive with accession number DRA001131 [13].

Prior to analysis by the proposed method, the mouse dendritic cell time-course RNA-Seq dataset was checked for read quality using FastQC [19]. The resulting reads for each of the three cell types were mapped to *M. musculus* mm10 genome RefSeq gene annotations using Bowtie1 [20] and Tophat2 [21]. Indices and annotations for Bowtie1 and Tophat2 were downloaded from the respective programs' websites. Per-base read quality scores and mapping rates for each sample are shown in S1 Fig. Reads that were successfully mapped by Tophat2 to the mouse transcriptome were then used to estimate the gene expressions in each time sample using Cufflinks [22]. Gene expression across different time samples in the same cell type were normalized as FPKM (fragment per kilobase of exon per million fragments mapped) and as a time series using Cuffdiff with option -T [22].

Before analyzing the processed RNA-Seq data, we first filtered out genes that have no expression or show limited changes in expression throughout the time series in all three cell types. We kept for subsequent analysis only those genes that in at least one of the cell types have a greater than 2-fold change between the maximum and minimum expressions, and have a maximum expression of greater than 5 fpkm. The remaining 5,676 genes were then z-normalized to mean of zero and variance of one.

**Caucasian RNA-Seq Data.** To test the robustness of the proposed method under different sample sizes, we have included RNA-Seq sequenced from mRNA obtained from the lymphoblastoid cell lines (LCL) of 60 Caucasian extended HapMap individuals [14]. The raw read count of each gene has been compiled and made available on the ReCount website (<http://bowtie-bio.sourceforge.net/recount/>) [15]. The raw read counts in a sample were normalized by dividing by the 75th percentile of the non-zero read counts of that sample [15, 23]. The total number of genes was further filtered down to 3,599 genes by requiring that the gene have at least one read in each of the 60 samples. Since these samples were not stimulated with LPS, they further test the sensitivity of the proposed method in datasets with weak signals.

## Gene expression profile similarity

To determine which genes are important to the transduction of signals between two distinct pathways, we make the assumption that genes with similar expression profiles are more likely to be interacting with each other. Given gene expression data from RNA-Seq experiments, one way of determining the similarity between two genes is to compute some measure of distance between their expression profiles. One of the most commonly used distance measures is the correlation distance, based on either the Spearman or the Pearson correlation coefficients. Non-correlation-based distance measures such as Euclidean, City-block, and Maximum distance, which are all special cases of Minkowski distances [24], have also been used as distance measures in clustering

algorithms for gene expressions [25]. For time-series gene expression data, Mahalanobis distance has been used as a distance measure to detect differential expression [26].

There is a vast amount of literature dealing with using distance measures to determine the similarity of two genes based on their gene expression. However, as with all distance-based approach, performance of each distance metric can be affected by the distribution and noise of the data. In particular, Euclidean distance does not consider correlation between data, whereas sample correlation coefficient is sensitive to outliers and sample size [27]. In this work, instead of using a distance measure and choosing a threshold for calling gene interaction, we propose the use of a sparse regression approach called elastic-net to predict gene interaction.

### Sparse regression

Current advances in sequencing technology have led to a tremendous growth in biomedical data, where sample dimensions, such as the number of genes or SNPs, have been growing at a much faster rate than that of number of samples. This phenomenon has emphasized the need for dimensionality reduction techniques in order to enhance the interpretability of statistical models used to analyze these data [28, 29]. One such example is lasso, a sparse regression model based on the  $\ell_1$  norm, that has been applied to many problems in bioinformatics and computational biology [30].

One disadvantage of lasso regression is that it selects at most  $T$  predictors with non-zero coefficients, where  $T$  is the number of samples [31]. In the case where there are  $P$  predictors and  $P \gg T$ , lasso regression may not be able to select enough predictors to model the dependent variable. Also, lasso may randomly select one variable from a group of variables with high pairwise correlations. In this work, we will use a sparse regression that overcomes these limitations by linearly combining the  $\ell_1$  and  $\ell_2$  penalties as regularization terms [31]. This sparse regression, or elastic-net, has the following optimization function:

$$\hat{\omega}^{EN} = \arg \min_{\omega} (\| \mathbf{y} - \mathbf{X}\omega \|^2 + (1 - \lambda) \| \omega \|^2 + \lambda \| \omega \|_1), \tag{1}$$

where  $\mathbf{y}$  is the dependent variable,  $\mathbf{X}$  is the independent variable or predictor,  $\omega$  is the coefficient of the predictor, and  $\lambda$  is a parameter that determines the sparsity of the model fitting.

In our work, genes  $a$  and  $b$  are defined as neighbor genes if they have “similar” expression profile. If gene  $a$  and gene  $b$  are neighbors and gene  $a$  belongs to pathway 1, then gene  $b$  is also considered as a neighbor of pathway 1. Let us first define  $\Omega$  as the set of all genes,  $\chi_1$  as the set of genes in pathway 1, and  $\chi_2$  as the set of genes in pathway 2, and there is no overlap between  $\chi_1$  and  $\chi_2$ , i.e.  $\chi_1 \cap \chi_2 = \emptyset$ . We can further define  $\Pi$  as the set of all genes not in pathway 1 and pathway 2, or  $\Pi = (\Omega \setminus \chi_1) \cap (\Omega \setminus \chi_2)$ , where  $\setminus$  denotes set difference. Then, our goal is to find the set of genes  $\Gamma_1 \in \Pi$  such that all genes in  $\Gamma_1$  are neighbors to pathway 1, and  $\Gamma_2 \in \Pi$  such that all genes in  $\Gamma_2$  are neighbors to pathway 2, and the set of shared neighbor genes between pathways 1 and 2 is denoted as  $\Gamma_{1 \cap 2}$ .

When using correlation distance, a pair of genes is considered to be associated if their correlation distance falls below the threshold. By computing the correlation distance of all genes in  $\Pi$  to  $\chi_1$  and to  $\chi_2$ , we can determine the neighbor gene sets  $\Gamma_1$  and  $\Gamma_2$ . In this work, we do not compute a distance measure and setting a threshold for forming edges between genes. Instead, we will use elastic-net regression to predict the association between genes in  $\Pi$  to gene  $a$  in pathway  $\chi$ , by forming an edge between gene  $a$  and those genes in  $\Pi$  that have non-zero coefficient after modeling the expression of  $a$  with the expressions of genes in  $\Pi$ .

To setup the problem, we define  $\mathbf{y}_i$  as the  $T \times 1$  expression profile vector of gene  $a_i$ ,  $a_i \in \chi$ , where  $T$  is the number of samples. Furthermore,  $\mathbf{X}$  is the  $T \times |\Pi|$  expression profile matrix of genes in  $\Pi$ , where  $|\Pi|$  is the number of genes in  $\Pi$ , and  $\omega_i$  is the  $P \times 1$  vector of regression

coefficients for the elastic-net regression model of  $a_i$ . With this formulation, we fit the expression profile gene  $a_i$  in  $\mathcal{X}$  using the sparse elastic-net regression, with genes in  $\Pi$  as predictors. The fitted  $P \times 1$  coefficient vector  $\widehat{\omega}_i^{EN}$  will have  $N \ll P$  nonzero coefficients, and the corresponding genes in  $\Pi$  are defined as neighbor genes to gene  $a_i$ .

To obtain all the neighbor genes of a pathway,

**Data:**  $\mathbf{X}$ : predictor gene expression matrix,  $\mathbf{y}_i$ : gene expression vector for gene  $a_i \in \mathcal{X}$   
**for**  $i = 1$  **to**  $|\mathcal{X}|$  **do**  
 1. Estimate  $\widehat{\omega}_i^{EN} = \arg \min_{\omega_i} (\|\mathbf{y}_i - \mathbf{X} \omega_i\|^2 + (1 - \lambda) \|\omega_i\|^2 + \lambda \|\omega_i\|_1)$ ;  
 2. Select the set of genes,  $\Gamma^i$ , with non-zero coefficients in  $\widehat{\omega}_i^{EN}$ ;  
**end**  
 Find  $\Gamma = (\cup_{i=1}^{|\mathcal{X}|} \Gamma^i)$ , the set of all neighbor genes to  $\mathcal{X}$ ;

**Algorithm 1:** Neighbor genes discovery

Then, to obtain the shared neighbor genes between pathways 1 and 2,

**Data:**  $\mathbf{X}$ : predictor gene expression matrix,  $\mathbf{y}_i$ : gene expression vector for gene  $a_i \in \mathcal{X}_1$ ,  $\mathbf{z}_j$ : gene expression vector for gene  $b_j \in \mathcal{X}_2$   
 Find  $\Gamma_1$ , the set of all neighbor genes to  $\mathcal{X}_1$ , using Algorithm 1;  
 Find  $\Gamma_2$ , the set of all neighbor genes to  $\mathcal{X}_2$ , using Algorithm 1;  
 Find the shared neighbor genes for pathways 1 and 2:  $\Gamma_{1 \cap 2} = \Gamma_1 \cap \Gamma_2$ ;

**Algorithm 2:** Shared neighbor genes discovery

Each predictor gene in  $\Gamma_{1 \cap 2}$  has non-zero coefficient for at least one gene in each of the two pathways. Since the gene expressions of these shared neighbor genes can be used to predict the expression of genes in the two pathways in the elastic-net regression model, we hypothesize that they are also good candidates as genes that link together the two pathways. Fig 1 illustrates the steps in Algorithm 1 for finding shared neighbor genes between pathways 1 and 2.

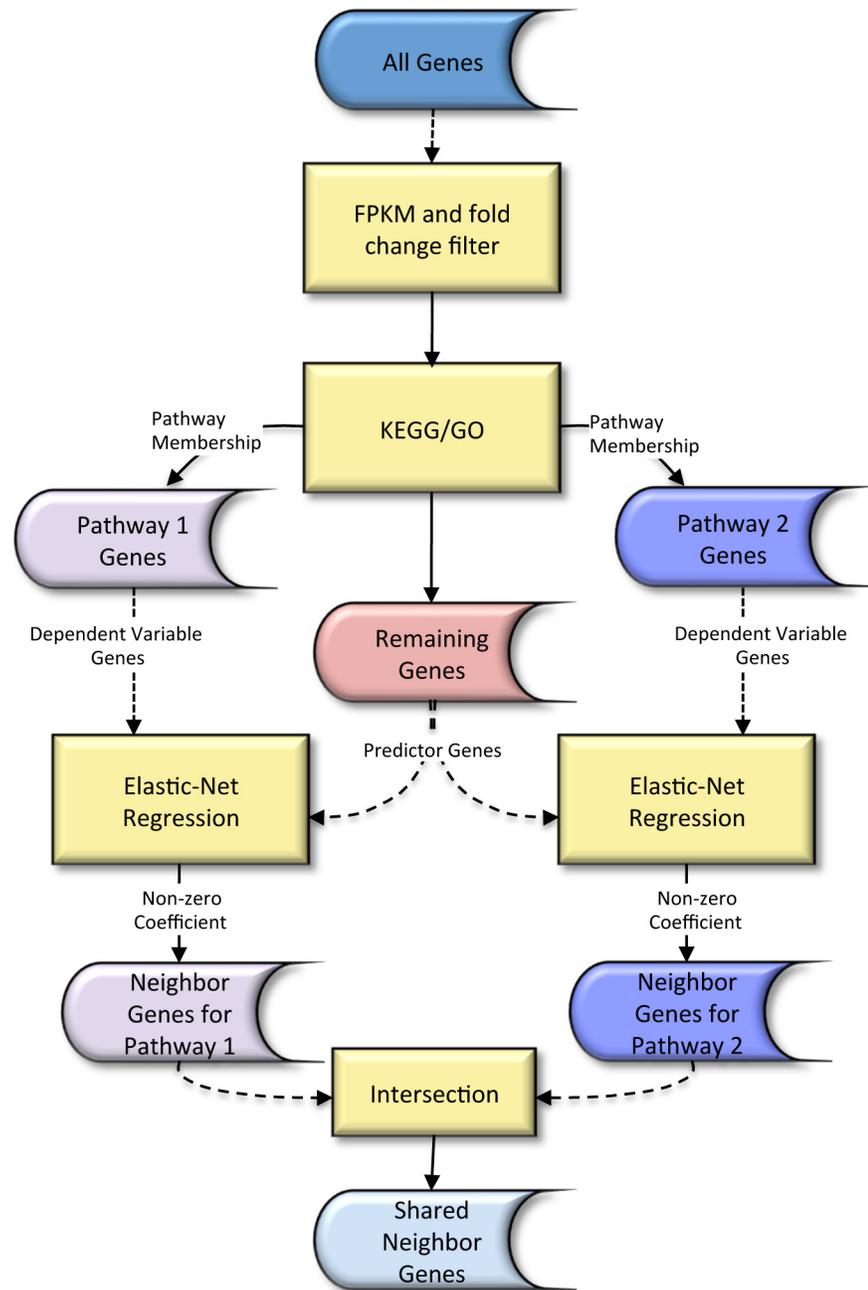
When applied to samples from different cell types, the shared neighbor genes found in different cell types for the same pathway pair may be different due to differential expressions. For example, in a gene knockout experiment, shared neighbor genes found only in the wild type sample but not the knockout sample are potentially paths between the two pathways that are affected by the gene knockout. To simplify our notation, we will drop the pathway subscripts and denote  $\Gamma_{W \setminus K}$  for some pathway pair as the set of shared neighbor genes found only in the wild type but not knocked out sample, and  $\Gamma_{W \cap K}$  as the set of shared neighbor genes that are found in both the wild type and the knocked out sample.

Through gene set enrichment analysis, we can find if any functional overrepresentation and statistical significance exist in the shared neighbor genes that exist under only a specific condition. In particular, we are interested in finding if certain shared neighbor genes exist only in the wild type sample, but not in the knockout sample. Such genes may have a role in connecting the two pathways and their functions are affected by the gene knockout in the upstream pathway. We will then compare the statistically significant overrepresented functions of these genes to known research findings to confirm the role of these genes in connecting the pathways.

## Results

### Mouse Dendritic Cell Knock-out Time-course Data

We applied the proposed method to a time-series mouse dendritic cell RNA-Seq experiment with wild type, MyD88 knockout, and TRIF knockout (KO) cell types. Since both of these



**Fig 1. Schematic diagram for finding shared neighbor genes between pathways 1 and 2.**

doi:10.1371/journal.pone.0137222.g001

adaptor proteins are key components of the TLR signaling pathway, we used the proposed method on each cell type to find shared neighbor genes between TLR signaling pathway and downstream pathways that are affected by immune response caused by LPS stimulation. We then found the set difference gene lists between the wild type and one of the knockout data to find neighbor genes that are unique to each cell type. For elastic-net regression, we used the glmnet implementation in R [32] with  $\lambda = 0.5$  and fitted the model to explain 75% of the

variance in each gene's time-series expression. For WGCNA, we set soft power to 8 for the adjacency computation.

From previous research, after knocking out MyD88 or TRIF in dendritic cells, we expected to observe changes in expressions in genes belonging to the downstream antigen processing and presentation, apoptosis, and Jak-Stat pathways [17, 33, 34]. We constructed the gene lists for these pathways by obtaining their gene lists from KEGG [1, 2] and AmiGO [35]. For each of the TLR-signaling—antigen processing and presentation, TLR-signaling—apoptosis, and TLR-signaling—Jak-Stat pathway pairs, we used genes not in the two pathways as predictor genes in elastic-net regression.

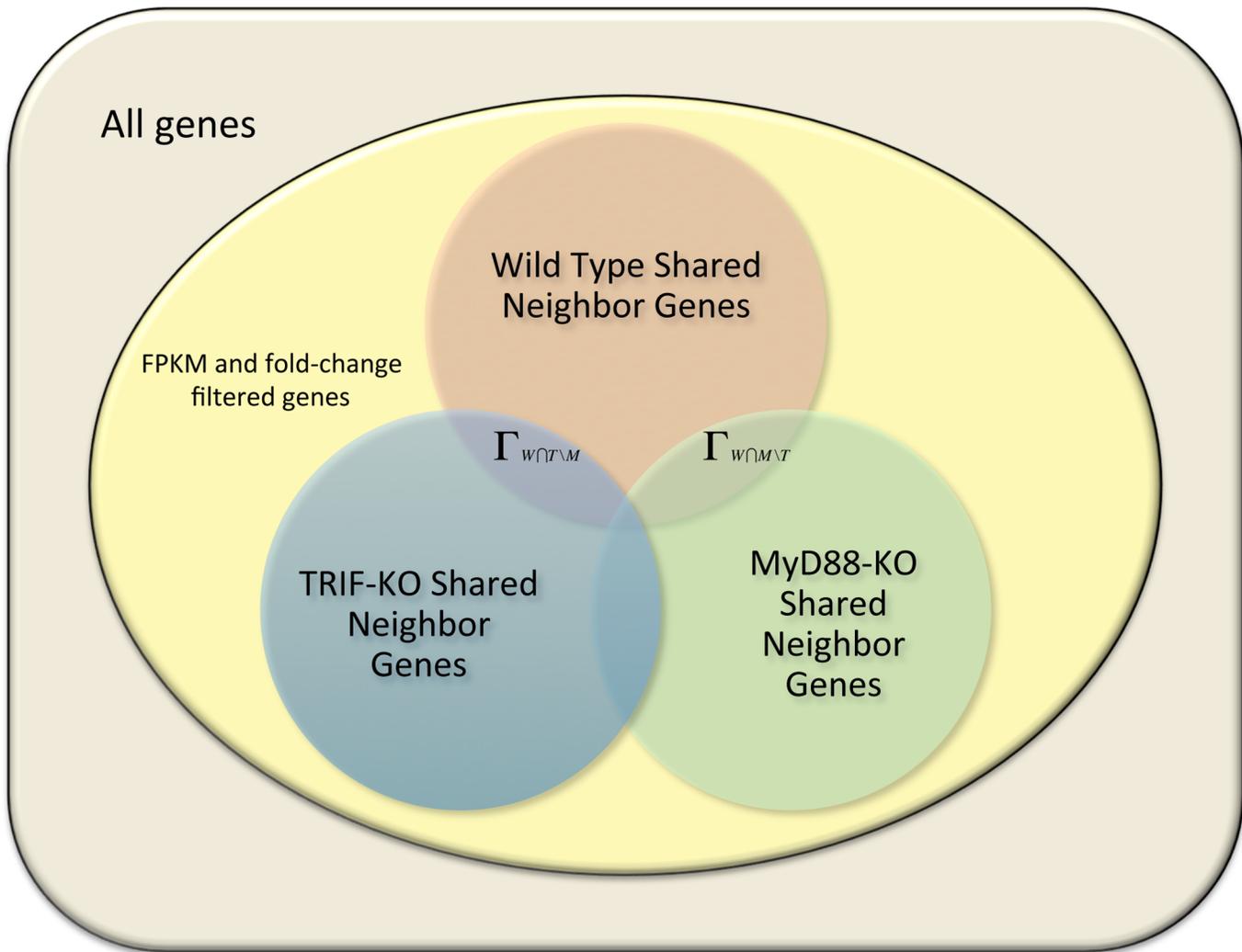
In the following discussion, we compared the shared neighbor genes in wild type to those in MyD88 KO and TRIF KO cell types to find potential links between TLR signaling pathway and a downstream pathway, and observed which of these links are affected by the gene knockouts. We denote  $\Gamma_{W \cap MYT}$  as shared neighbor genes that exist between two pathways in both the wild type and MyD88-KO samples, but not in the TRIF-KO sample. This set represents those genes that are associated with the TRIF-dependent part of the TLR-signaling pathway, and their functions are affected by the TRIF-KO. Similarly,  $\Gamma_{W \cap TM}$  denotes those shared neighbor genes that exist in both the wild type and the TRIF-KO samples, but not the MyD88-KO sample. Fig 2 shows a Venn diagram of the two gene sets. In our comparisons we used Weighted Correlation Network Analysis (WGCNA) as well as a correlation distance implementation using  $\mu = 0.1$  as the threshold for finding neighbor genes. We then used gene ontology overrepresentation analysis on these gene set difference lists to find important functions between the pathways that are disabled by the knockout, which knocked out path are these functions associated with, and confirmed our findings through prior research.

**Antigen processing and presentation.** Dendritic cells are critical to the adaptive immune mechanism by acting as an initiator for activating T cells and initiating primary and memory immune responses [36], and presentation of pathogens is accomplished through major histocompatibility complex (MHC) class I or MHC class II molecules. It is well known that when dendritic cells are stimulated by LPS, the downstream antigen processing and presentation functions are activated through the upstream TLR-signaling pathway [37].

We used the proposed method to find shared neighbor genes between TLR signaling pathway and antigen processing and presentation for each of the wild type, MyD88 knockout, and TRIF knockout cell types, denoted as  $\Gamma_W$ ,  $\Gamma_M$ , and  $\Gamma_T$ , respectively. We then constructed the sets  $\Gamma_{W \cap TM}$  and  $\Gamma_{W \cap MYT}$  as described in the Materials and Methods Section. The same steps were followed for correlation distance and WGCNA. The corresponding set difference gene sets using correlation distances are denoted as  $\Theta_{W \cap TM}$  and  $\Theta_{W \cap MYT}$ , and those for WGCNA are denoted as  $\Phi_{W \cap TM}$  and  $\Phi_{W \cap MYT}$ .

In Figs 3 and 4 we have drawn figures of shared neighbor genes with their connected pathway genes in  $\Gamma_{W \cap TM}$  and  $\Gamma_{W \cap MYT}$  using Cytoscape [38], respectively.

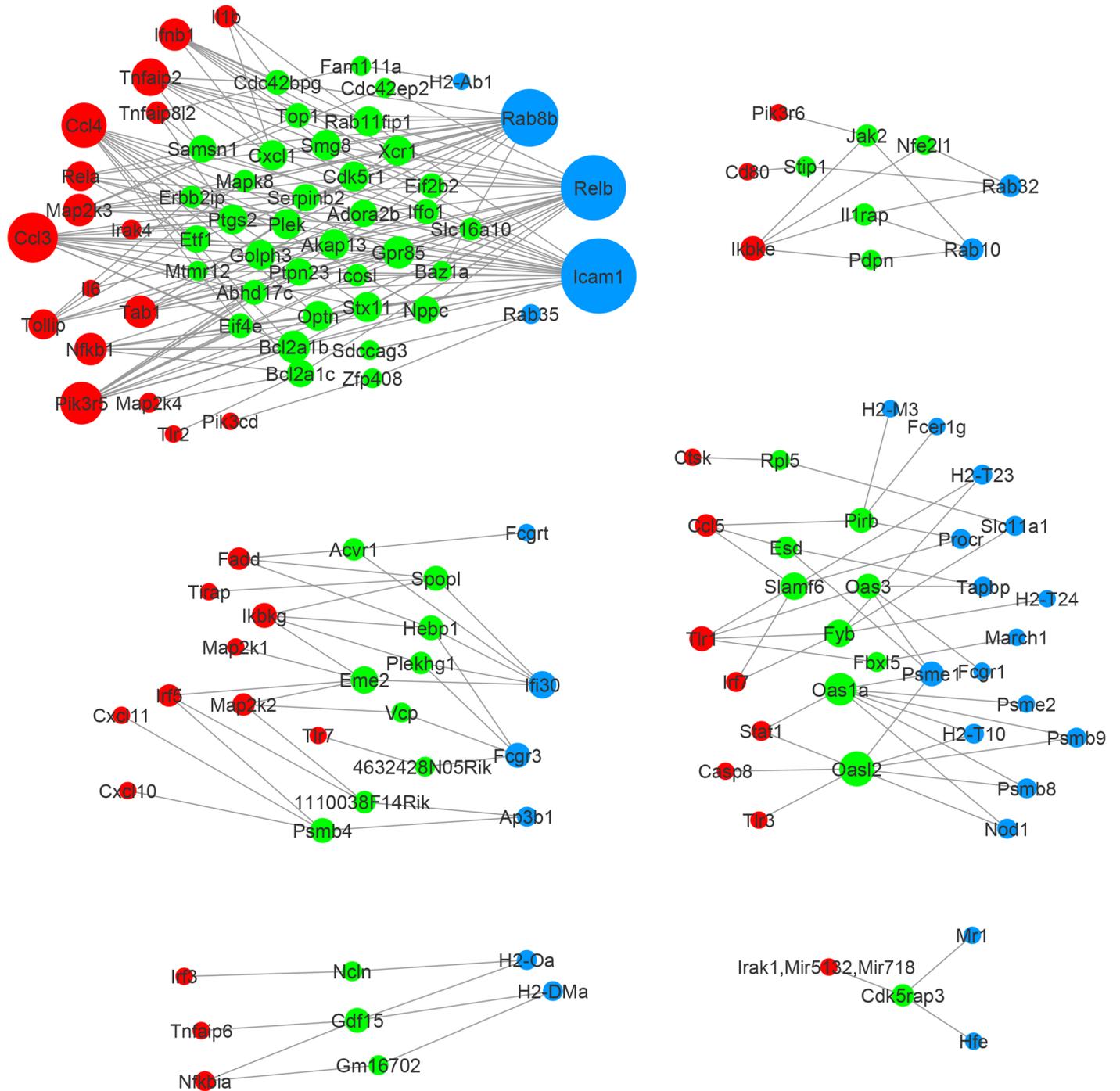
Genes in each of the six sets can be found in Table A in S1 File. For sets  $\Gamma_{W \cap TM}$  and  $\Gamma_{W \cap MYT}$  we found 63 and 108 genes, respectively. For sets  $\Theta_{W \cap TM}$  and  $\Theta_{W \cap MYT}$ , we found 689 and 276 genes, respectively. For sets  $\Phi_{W \cap TM}$  and  $\Phi_{W \cap MYT}$ , we found 186 and 285 genes, respectively. The numbers of genes in the two gene sets found by elastic-net are much closer to each other than those found by the correlation distance implementation, where the set with more genes is almost 3 times larger than the smaller one. Compared to the correlation distance implementation, the two gene sets found by WGCNA also have more comparable numbers of genes. This is an advantage for the elastic-net method and WGCNA due to the impact that gene list size has on the significance of terms discovered, making the overrepresentation results for the elastic-net-generated lists much more comparable.



**Fig 2. Venn diagram of shared neighbor genes from wild type, MyD88-KO, and TRIF-KO samples.**

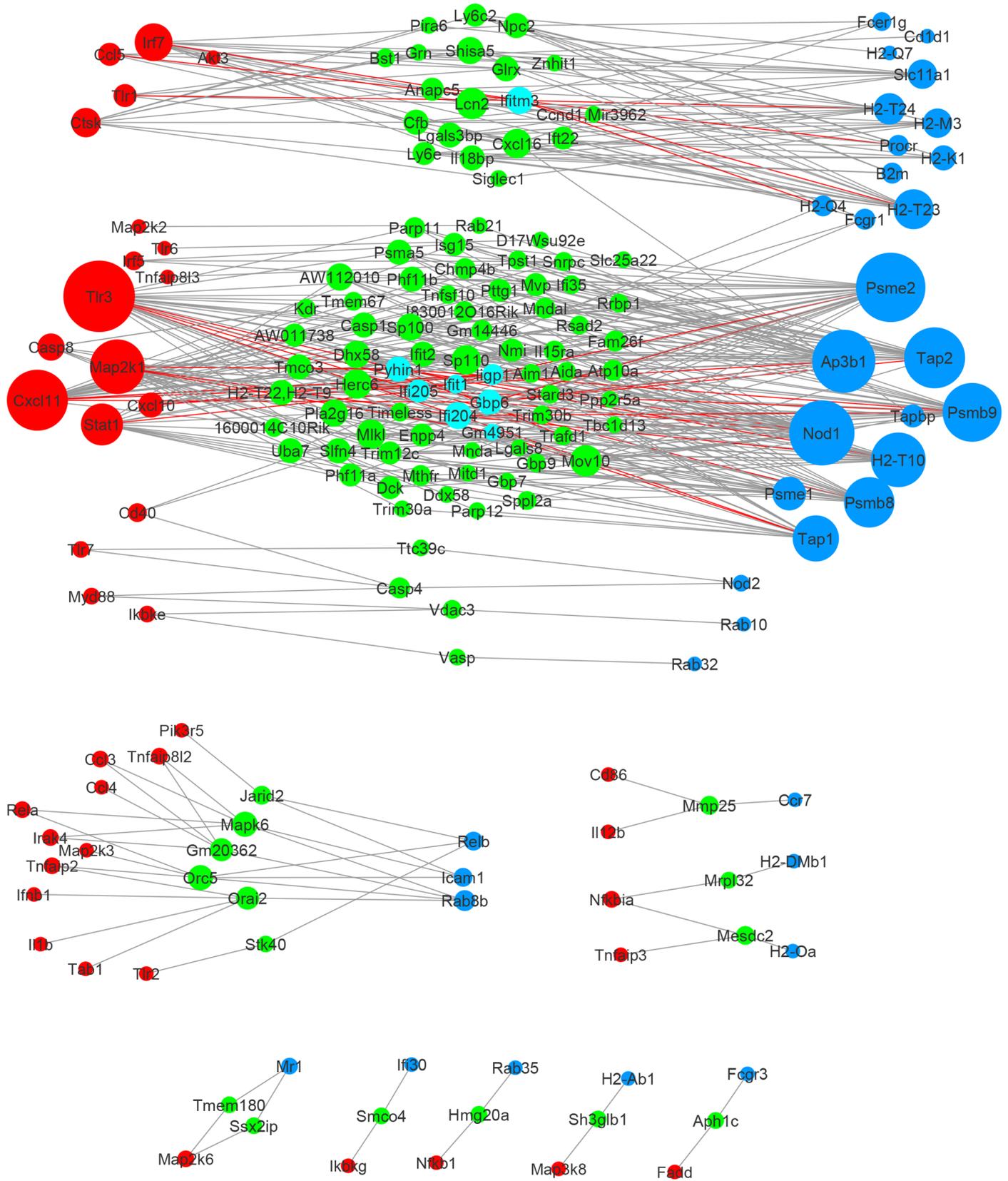
doi:10.1371/journal.pone.0137222.g002

To find overrepresented GO terms in the gene lists we used Gorilla [39] with  $\Pi$  as background. The significant gene ontology terms and the genes with those GO terms in  $\Gamma_{W \cap TM}$  and  $\Gamma_{W \cap MT}$  are listed in Tables 1 and 2, respectively. The top 5 significant gene ontology terms and their FDR for elastic-net, correlation distance, and WGCNA are shown in Table B in S1 File. Of the 6 gene sets, only  $\Gamma_{W \cap MT}$  and  $\Phi_{W \cap MT}$  contain statistically significant gene ontology terms related to immune process after multiple testings correction using FDR. In particular, we found that both  $\Gamma_{W \cap MT}$  and  $\Phi_{W \cap MT}$  are significantly represented with genes in response to cytokine (FDR = 3.63E-4 for elastic-net and FDR = 1.19E-9 WGCNA) and response to interferon- $\beta$  (FDR = 1.39E-4 for elastic-net and FDR = 2.49E-9 for WGCNA). Both of these gene sets are overrepresented in similar GO terms, with WGCNA having the more significant gene set in terms of FDR, but the elastic-net method also having significant FDR. On the other hand, for the correlation distance approach, while  $\Theta_{W \cap MT}$  also contains genes with the gene ontology term response to interferon- $\beta$ , its FDR is 1E0, with no other terms showing any statistical significance.



**Fig 3. Network edges of shared neighbor genes found in wildtype and TRIF-KO, but not in MyD88-KO, by elastic-net.** Red nodes: TLR signaling pathway genes. Blue nodes: Antigen processing and presentation genes. Green nodes: Shared neighbor genes.

doi:10.1371/journal.pone.0137222.g003



**Fig 4. Network edges of shared neighbor genes found in wildtype and MyD88-KO, but not in TRIF-KO, by elastic-net.** Red nodes: TLR signaling pathway genes. Blue nodes: Antigen processing and presentation genes. Green nodes: Shared neighbor genes. Cyan nodes: Shared neighbor genes with "reponse to interferon- $\beta$ " GO term. Red edges: edges from cyan genes to pathway genes.

doi:10.1371/journal.pone.0137222.g004

**Table 1. Significant GO terms and genes for  $\Gamma_{W \cap TM}$  between TLR Signaling Pathway and Antigen Processing and Presentation.**

GO Terms	Genes
positive regulation of determination of dorsal identity	Acvr1, Mapk8
positive regulation of response to wounding	Jak2, Ptgs2, Adcra2b, Plek, Pdpn
positive regulation of phosphatase activity	Jak2, Rpl5, Plek

doi:10.1371/journal.pone.0137222.t001

**Table 2. Significant GO terms and genes for  $\Gamma_{W \cap MT}$  between TLR Signaling Pathway and Antigen Processing and Presentation.**

GO Terms	Genes
response to interferon- $\beta$	Ifi205, Ifit1, Gm4951, Gbp6, Ifitm3, Ifi204, Pyhin1, ligp1
cellular response to interferon- $\beta$	Ifi205, Ifit1, Gbp6, Gm4951, Ifi204, Pyhin1, ligp1
response to cytokine	Gbp7, Gbp6, Gm4951, Pyhin1, Ifit2, Ifi205, Ifit1, Isg15, Ifi204, Ifitm3, Gbp9, lig1, Cxcl16
response to other organism	Gbp7, Sp110, Gbp6, Casp1, Ddx58, Dhx58, Ifit2, Ifit1, Isg15, Ifitm3, Gbp9, ligp1, Rsad2, Lcn2
cellular response to cytokine stimulus	Ifi205, Gbp7, Ifit1, Gm4951, Gbp6, Ifi204, Gbp9, Pyhin1, lig1, Ifit2

doi:10.1371/journal.pone.0137222.t002

Through literature search, we were able to confirm the findings by elastic-net and WGCNA with prior study showing that TRIF is responsible for the induction of interferon- $\beta$  [40], and more recent research showing that the inhibition of interferon- $\beta$  impairs the antigen presentation functions of dendritic cells [41]. Using the proposed methodology, we were able to predict correctly that genes related to the response to interferon- $\beta$  are acting as intermediary between TLR signaling pathway and genes responsible for antigen processing and presentation.

**Apoptosis.** Apoptosis of dendritic cells plays an important role in the balancing of immune tolerance and the development of autoimmunity [42]. An excessive activation of dendritic cells can induce tissue-specific and systemic autoimmune symptoms [42–44], and an environment where significant dendritic cell apoptosis occurs is immunosuppressive [45]. It has been shown that dendritic cell apoptosis is controlled by the TLR4-mediated TRIF-dependent signaling pathway [46] and that Type-1 interferons are necessary and sufficient for the induction of apoptosis for dendritic cells [47]. In particular, interferon- $\gamma$  has been shown to induce nitric oxide synthase in mouse dendritic cells, and the production of nitric oxide is associated with dendritic cell apoptosis [48]. Furthermore, cytokines has been shown to be the path of TRIF-induced apoptosis [49].

For TLR signaling pathway and apoptosis we have found 65 and 63 genes for  $\Gamma_{W \cap TM}$  and  $\Gamma_{W \cap MT}$ , respectively. For  $\Theta_{W \cap TM}$  and  $\Theta_{W \cap MT}$  we found 842 and 337 genes, respectively. For

**Table 3. Significant GO terms and genes for  $\Gamma_{W \cap T \cap M}$  between TLR Signaling Pathway and Apoptosis.**

GO Terms	Genes
protein dephosphorylation	Ptpre, Ppp1r15b, Ptpn23, Dusp16, Fbxw11, Ssh2
dephosphorylation	Ptpre, Ppp1r15b, Ptpn23, Dusp16, Fbxw11, Ssh2, Mtmr14
ubiquitin-dep. protein catabolic proc.	Ptpn23, Usp16, Siah2, Psmb7, Rffl, Ube4a, Fbxw11
proteolysis	Bcl10, Siah2, Ptpn23, Usp16, Psmb7 Rffl, Ndel1, Plat, Ube4a, Fbxw11, Metap1
positive reg. of platelet activation	Plek, Pdpn

doi:10.1371/journal.pone.0137222.t003

$\Phi_{W \cap T \cap M}$  and  $\Phi_{W \cap M \cap T}$  we found 201 and 274 genes, respectively. The genes for each gene set are listed in Table A in [S1 File](#). Again, the numbers of genes found are much more comparable in the elastic-net implementation and WGCNA, with elastic-net finding almost exactly the same number of genes in the two shared neighbor gene sets.

We analyzed the gene sets for overrepresented GO terms, and the top 5 significant gene ontology terms for each are shown in Table C in [S1 File](#). The significant gene ontology terms and the genes with those GO terms for  $\Gamma_{W \cap T \cap M}$  and  $\Gamma_{W \cap M \cap T}$  are listed in Tables 3 and 4, respectively. For  $\Gamma_{W \cap M \cap T}$ , we found that response to interferon- $\gamma$  is significantly overrepresented with FDR = 1.15E-2, and no immune related terms were found to be significant in the  $\Gamma_{W \cap T \cap M}$  set. Response to interferon- $\gamma$  was also found to be significant in  $\Phi_{W \cap M \cap T}$  with FDR = 8.75E-4. Again, both elastic-net and WGCNA found similar overrepresented GO terms, with WGCNA having more significant FDR values.

For correlation distance, while  $\Theta_{W \cap M \cap T}$  contains several highly significant immune related GO terms, it is significantly overrepresented only in response to interferon- $\beta$  (FDR = 1.65E-4), but not interferon- $\gamma$ . In this case, both the elastic-net implementation and WGCNA were again able to find result that is supported by existing studies. While the correlation distance implementation was also able to find that significant immune related processes are altered by TRIF-KO, it's results are not as precise and accurate.

**Jak-Stat signaling pathway.** The Jak-Stat signaling pathway is responsible for signal transduction in development and homeostasis in animals, and is the primary signal mechanism for cytokines [50]. It has been pointed out that because the Jak-Stat signaling pathway is downstream of interferon- $\beta$  production, it should be affected by the TRIF-dependent pathway [51], but not the MyD88-dependent pathway [52].

The genes in  $\Gamma_{W \cap T \cap M}$ ,  $\Gamma_{W \cap M \cap T}$ ,  $\Theta_{W \cap T \cap M}$ , and  $\Theta_{W \cap M \cap T}$  are listed in Table A in [S1 File](#). For sets  $\Gamma_{W \cap T \cap M}$  and  $\Gamma_{W \cap M \cap T}$  we have found 137 and 114 genes, respectively. For sets  $\Theta_{W \cap T \cap M}$  and  $\Theta_{W \cap M \cap T}$  we found 734 and 308 genes, respectively. For sets  $\Phi_{W \cap T \cap M}$  and  $\Phi_{W \cap M \cap T}$  we found 378 and 407 genes, respectively.

**Table 4. Significant GO terms and genes for  $\Gamma_{W \cap M \cap T}$  between TLR Signaling Pathway and Apoptosis.**

GO Terms	Genes
cellular response to cytokine stimulus	Gbp7, Gm4951, Gbp6, Ifi204, Gbp9 Gbp3, ligp1, Ifit2
response to protozoan	Gbp7, Gbp6, Gbp9, Gbp3, ligp1
cellular response to interferon- $\beta$	Gbp6, Gm4951, Ifi204, Gbp3, ligp1
cellular response to interferon- $\gamma$	Gbp7, Gbp6, Gbp9, Gbp3

doi:10.1371/journal.pone.0137222.t004

**Table 5. Significant GO terms and genes for  $\Gamma_{W \cap TMM}$  between TLR Signaling and Jak-Stat Pathways.**

GO Terms	Genes
positive regulation of cell cycle	Cdk5r1, Bcl2l11, Vps4b, Hspa2, Igf1r Tnf, E2f8, Eif4e, Nr4a3, Mad21
NIK/NF- $\kappa$ B signaling	Nfkb2, P49/p100, Relb, Rel

doi:10.1371/journal.pone.0137222.t005

The top significant gene ontology terms for each gene set are shown in Table D in [S1 File](#). The significant gene ontology terms and the genes with those GO terms for  $\Gamma_{W \cap TMM}$  and  $\Gamma_{W \cap MT}$  are listed in Tables 5 and 6, respectively. For TLR signaling Jak-Stat pathways, both the elastic-net implementation and WGCNA discovered in  $\Gamma_{W \cap MT}$  and in  $\Phi_{W \cap MT}$  genes that are overrepresented with the cellular response to interferon- $\beta$  term (FDR = 3.48E-5 and 9.56E-9, respectively), and found no significant functional overrepresentation related to immune process present in  $\Gamma_{W \cap TMM}$  nor  $\Phi_{W \cap TMM}$ , in agreement with existing literature. On the other hand, the gene sets obtained from correlation distance implementation also did not find any GO terms with significant functional overrepresentation. For  $\Gamma_{W \cap TMM}$ , gene ontology analysis did find several terms relating to antigen processing and presentation, but the lowest FDR of these terms is only 6.16E-1.

### Caucasian RNA-Seq Data

With the Caucasian RNA-Seq datasets, we again applied the elastic-net, correlation distance, and WGCNA to discover shared neighbor genes between TLR signaling pathway and antigen processing and presentation, apoptosis, and Jak-Stat pathways. We used the same parameters as used in the mouse dendritic cells analysis. Since there are no gene knockouts in this experiment, we only compared the shared neighbor gene sets found by elastic-net, correlation distance, and WGCNA in the wild type samples, denoted as  $\Gamma$ ,  $\Theta$ , and  $\Phi$ , respectively. The gene sets are listed in Table A in [S2 File](#), and the top overrepresented GO terms for shared neighbor genes between TLR signaling pathway and antigen processing and presentation, apoptosis, and Jak-Stat pathways are listed in Tables B, C, and D in [S2 File](#), respectively.

Elastic-net, correlation distance, and WGCNA found 82, 42, and 173 shared neighbor genes between TLR signaling pathway and antigen processing and presentation. In Table B in [S2 File](#), we can see that elastic net found that type I interferon signaling pathway and response to type I interferon to be significantly overrepresented with p-values 4.20E-4 and 5.53E-4, respectively. While specific terms related to interferon- $\beta$  was not found to be overrepresented, it is a member of the human type I interferon family. On the other hand, for correlation distance and

**Table 6. Significant GO terms and genes for  $\Gamma_{W \cap MT}$  between TLR Signaling and Jak-Stat Pathways.**

GO Terms	Genes
cellular response to interferon- $\beta$	lfi205, lfit1, Gbp6, Gm4951, lfi204 Gbp3, ligp1, Gbp2
response to cytokine	lfi205, Gbp7, lfit1, Gm4951, Gbp6, lsg1, lfi204, Trim56, Ptg2, Gbp3 Gbp3, ligp1, Gbp2
multi-organism process	Sp110, Gbp7, Gbp6, Chmp4b, Trim56, Hdac1 lfi1, Gbp3, Ddx58, Ccnk, Gbp2, Rsad2 Tap2, lfit1, lsg15, Ddx60, ligp1
defense response to other organism	Gbp7, lfit1, Gbp6, lsg15, Trim56, lfi1 Ddx60, Gbp3, ligp1, Ddx58, Gbp2, Rsad2

doi:10.1371/journal.pone.0137222.t006

WGCNA, no immune process related GO terms were found to be overrepresented. The top terms for both are populated with metabolic and biosynthetic processes terms.

For apoptosis, 110, 142, and 794 shared neighbor genes were found by elastic-net, correlation distance, and WGCNA, respectively. From GO analysis, we can see that correlation distance and WGCNA found no overrepresented immune related GO terms, while elastic-net found lymphocyte and leukocyte mediated immunity to be overrepresented, which include genes such as FCER2 (CD23), RasGRP1, and HLA-E. From literature we know that CD23 is induced by TLR4 [53], and that high apoptotic rates are often correlated to the expression of CD23 [54]. RasGRP1 is a guanine nucleotide exchange factor whose expression is upregulated by LPS and other TLR agonists [55], and promotes B cell receptor-induced apoptosis [56]. Furthermore, the transcription of HLA-E, the human major histocompatibility complex (MHC) class Ib gene, is shown to be mediated by interferon- $\gamma$  [57], and is known to elicit apoptosis in natural killer cells [58].

For Jak-Stat pathway, 122, 95, and 268 shared neighbor genes were found by elastic-net, correlation distance, and WGCNA, respectively. In this case, we see that none of the shared neighbor gene sets are overrepresented in immune-related GO terms. Shared neighbor gene set found by elastic-net ( $\Gamma$ ) is enriched in phosphorylation and metabolic processes, whereas  $\Theta$  and  $\Phi$  are enriched with GO terms related to transcription and RNA synthesis.

These results show that with the Caucasian RNA-Seq data, the proposed elastic-net approach is more sensitive than WGCNA and correlation distance. Since these Caucasian individual samples had not been stimulated with LPS, the immune response pathways are not expected to be consistently activated across the samples. Therefore, any relationship the upstream TLR signaling pathway has with a downstream pathway, and with any intermediary genes will be weak, making the inference of gene-gene interactions a difficult task. In this regard, while the significances are not high after correcting for multiple testing, several immune process related GO terms with significant p-values were found by elastic-net for antigen processing and presentation and for apoptosis. In the case of apoptosis, several genes are shown by previous research to be related to both the TLR signaling pathway and apoptosis, despite the fact that the GO terms are not known to be their intermediary. These results show that elastic-net can be more sensitive in detecting gene-to-gene relationships when the signals are weak.

## Conclusion

The discovery of genes that link together the activities of different pathways are crucial to the advance of pathway analysis and systems biology. Here we have proposed a methodology using sparse regression approach to specifically discover genes that are shared neighbors of two different pathways. We have also shown that the method chosen to select neighbor genes can greatly affect the outcome of the subsequent overrepresentation analysis. In this paper, we have adopted the elastic-net regression approach and select those predictor genes with non-zero coefficients as neighbors to the gene whose expression was modeled. We show in a series of comparisons that the elastic-net implementation can predict gene-to-gene relationships enriched with comparable GO terms to those predicted by WGCNA. At the same time, elastic-net is shown to be more sensitive than WGCNA in datasets that contains only weak relationships between genes, and is superior to a simple correlation distance implementation in all cases tested.

It should be noted that while it is possible that a gene in one pathway studied in our work may interact directly with a gene in the other pathway, we only attempt to find genes outside of the two pathways, and whose expression profiles show interactions with both pathways. In order to find direct interactions, when modeling genes in pathway 1, we can add genes from

pathway 2 to the rest of the genes, and those genes in pathway 2 with non-zero coefficients can then be linked directly to pathway 1, and vice versa. So far we have used only the non-zero coefficient requirement as a selection criteria for neighbor genes of those in the two pathways. One future direction for the development of this method is to introduce a more sophisticated selection criteria to more precisely select neighbor genes. One such approach can be the use of the coefficient values as a criteria, where we can rank the predictor genes with non-zero coefficients and select only the top genes as neighbors. Or we could compute the geometric mean of the number of times a predictor is non-zero for the two pathways. Furthermore, the gene selection can be further validated by having p-values for the fitted coefficients in elastic-net. While currently there is no known approach for computing the p-value for the coefficients in elastic-net, one may try to do a large number of permutations of the gene expressions of the dependent variable gene, and see how often a selected gene from the original is selected in the permutations.

## Supporting Information

**S1 File. Significant genes and GO terms found by elastic-net, WGCNA and correlation distance for time-course mouse dendritic cell RNA-Seq data.** Complete gene lists of  $\Gamma_{W \cap TM}$ ,  $\Gamma_{W \cap MT}$ ,  $\Theta_{W \cap TM}$ ,  $\Theta_{W \cap MT}$ ,  $\Phi_{W \cap TM}$ , and  $\Phi_{W \cap MT}$  (Table A). Top GO terms for shared neighbor genes between TLR-signaling pathway and antigen processing and presentation genes (Table B). Top GO terms for shared neighbor genes between TLR-signaling pathway and apoptosis genes (Table C). Top GO terms for shared neighbor genes between TLR-signaling pathway and Jak-Stat pathway genes (Table D). (XLSX)

**S2 File. Significant genes and GO terms found by elastic-net, WGCNA and correlation distance for HapMap Caucasian individual RNA-Seq data.** Complete gene lists of  $\Gamma$ ,  $\Theta$ , and  $\Phi$  (Table A). Top GO terms for shared neighbor genes between TLR-Signaling pathway and antigen processing and presentation genes (Table B). Top GO terms for shared neighbor genes between TLR-Signaling pathway and apoptosis genes (Table C). Top GO terms for shared neighbor genes between TLR-Signaling pathway and Jak-Stat pathway genes (Table D). (XLSX)

**S1 Fig. Per-base quality scores and mapping-rate for all samples.** (PDF)

## Acknowledgments

Computational resources were provided by the supercomputer system at Human Genome Center, the Institute of Medical Science, the University of Tokyo.

## Author Contributions

Conceived and designed the experiments: KL AP KN. Performed the experiments: KL. Analyzed the data: KL AP KN. Wrote the paper: KL AP. Provided computational structure: KN.

## References

1. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000; 28(1):27–30. doi: [10.1093/nar/28.1.27](https://doi.org/10.1093/nar/28.1.27) PMID: [10592173](https://pubmed.ncbi.nlm.nih.gov/10592173/)
2. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 2013; 42(D1):D199–D205. doi: [10.1093/nar/gkt1076](https://doi.org/10.1093/nar/gkt1076) PMID: [24214961](https://pubmed.ncbi.nlm.nih.gov/24214961/)

3. Milacic M, Haw R, Rothfels K, Wu G, Croft D, Hermjakob H, et al. Annotating cancer variants and anti-cancer therapeutics in Reactome. *Cancers*. 2012; 4(4):1180–1211. doi: [10.3390/cancers4041180](https://doi.org/10.3390/cancers4041180) PMID: [24213504](https://pubmed.ncbi.nlm.nih.gov/24213504/)
4. Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, et al. The Reactome pathway knowledgebase. *Nucleic Acids Res*. 2013; 42(D1):D472–D477. doi: [10.1093/nar/gkt1102](https://doi.org/10.1093/nar/gkt1102) PMID: [24243840](https://pubmed.ncbi.nlm.nih.gov/24243840/)
5. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 2013; 41(D1):D808–D815. doi: [10.1093/nar/gks1094](https://doi.org/10.1093/nar/gks1094) PMID: [23203871](https://pubmed.ncbi.nlm.nih.gov/23203871/)
6. Chatr-aryamontri A, Breitkreutz BJ, Heinicke S, Boucher L, Winter A, Stark C, et al. The BioGRID interaction database: 2013 update. *Nucleic Acids Res*. 2013; 41(D1):D816–D823. doi: [10.1093/nar/gks1158](https://doi.org/10.1093/nar/gks1158) PMID: [23203989](https://pubmed.ncbi.nlm.nih.gov/23203989/)
7. Khatri P, Sirota M, Butte AJ. Ten Years of Pathway Analysis: Current Approaches and Outstanding Challenges. *PLoS Comput Biol*. 2012; 8(2):e1002375. doi: [10.1371/journal.pcbi.1002375](https://doi.org/10.1371/journal.pcbi.1002375) PMID: [22383865](https://pubmed.ncbi.nlm.nih.gov/22383865/)
8. Donato M, Xu Z, Tomoiaga A, Granneman JG, MacKenzie RG, Bao R, et al. Analysis and correction of crosstalk effects in pathway analysis. *Genome Res*. 2013; 23:1885–1893. doi: [10.1101/gr.153551.112](https://doi.org/10.1101/gr.153551.112) PMID: [23934932](https://pubmed.ncbi.nlm.nih.gov/23934932/)
9. Hu R, Qiu X, Glazko G, Klebanov L, Yakovlev A. Detecting intervene correlation changes in microarray analysis: a new approach to gene selection. *BMC Bioinformatics*. 2009; 10(20): doi: [10.1186/1471-2105-10-20](https://doi.org/10.1186/1471-2105-10-20)
10. Mentzen WI, Floris M, de la Fuente A. Dissecting the dynamics of dysregulation of cellular processes in mouse mammary gland tumor. *BMC Genomics*. 2009; 10(601). doi: [10.1186/1471-2164-10-601](https://doi.org/10.1186/1471-2164-10-601) PMID: [20003387](https://pubmed.ncbi.nlm.nih.gov/20003387/)
11. Leonardson AS, Zhu J, Chen Y, Wan K, Lamb JR, Reitman M, et al. The effect of food intake on gene expression in human peripheral blood. *Hum Mol Genet*. 2009; 19(1):159–169. doi: [10.1093/hmg/ddp476](https://doi.org/10.1093/hmg/ddp476)
12. Jung S, Kim S. EDDY: a novel statistical gene set test method to detect differential genetic dependencies. *Nucleic Acid Res*. 2014; 24(7):e60. doi: [10.1093/nar/gku099](https://doi.org/10.1093/nar/gku099)
13. Patil A, Kumagai Y, Liang K, Suzuki Y, Nakai K. Linking transcriptional changes over time in stimulated dendritic cells to identify gene networks activated during the innate immune response. *PLoS Comput Biol*. 2013; 9(11). doi: [10.1371/journal.pcbi.1003323](https://doi.org/10.1371/journal.pcbi.1003323)
14. Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, et al. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature*. 2010; 464(7289):773–777. doi: [10.1038/nature08903](https://doi.org/10.1038/nature08903) PMID: [20220756](https://pubmed.ncbi.nlm.nih.gov/20220756/)
15. Frazee AC, Langmead B, Leek JT. ReCount: A multi-experiment resource of analysis-ready RNA-seq gene count datasets. *BMC Bioinformatics*. 2011; 12(449). doi: [10.1186/1471-2105-12-449](https://doi.org/10.1186/1471-2105-12-449) PMID: [22087737](https://pubmed.ncbi.nlm.nih.gov/22087737/)
16. Takeda K, Akira S. TLR signaling pathways. *Semin Immunol*. 2004; 16(1):3–9. doi: [10.1016/j.smim.2003.10.003](https://doi.org/10.1016/j.smim.2003.10.003) PMID: [14751757](https://pubmed.ncbi.nlm.nih.gov/14751757/)
17. Kawai T, Takeuchi O, Fujita T, Inoue J, Muhlrath PF, Sato S, et al. Lipopolysaccharide stimulates the MyD88-independent pathway and results in activation of IFN-regulatory factor 3 and the expression of a subset of lipopolysaccharide-inducible genes. *J Immunol*. 2001; 167(10):5587–5894. doi: [10.4049/jimmunol.167.10.5887](https://doi.org/10.4049/jimmunol.167.10.5887)
18. Reim D, Rossmann-Bloek T, Jusek G, Prazeres da Costa O, Holzmann B. Improved host defense against septic peritonitis in mice lacking MyD88 and TRIF is linked to a normal interferon response. *J Leuk Biol*. 2011; 90(3):613–620. doi: [10.1189/jlb.1110602](https://doi.org/10.1189/jlb.1110602)
19. Andrew S. FastQC: A quality control tool for high throughput sequence data;. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
20. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009; 10(R25): doi: [10.1186/gb-2009-10-3-r25](https://doi.org/10.1186/gb-2009-10-3-r25)
21. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013; 14(R36): doi: [10.1186/gb-2013-14-4-r36](https://doi.org/10.1186/gb-2013-14-4-r36)
22. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010; 28:511–515. doi: [10.1038/nbt.1621](https://doi.org/10.1038/nbt.1621) PMID: [20436464](https://pubmed.ncbi.nlm.nih.gov/20436464/)
23. Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics*. 2010; 11(94). doi: [10.1186/1471-2105-11-94](https://doi.org/10.1186/1471-2105-11-94) PMID: [20167110](https://pubmed.ncbi.nlm.nih.gov/20167110/)

24. Lee MLT. Analysis of Microarray Gene Expression Data. Springer Science and Business Media. Springer New York Inc.; 2004.
25. Hartuv E, Schmitt A, Lange J, Meier-Ewert S, Lehrach H, Shamir R. An Algorithm for Clustering cDNAs for Gene Expression Analysis. In: Proceedings of the Third Annual International Conference on Computational Molecular Biology. RECOMB'99. New York, NY, USA: ACM; 1999. p. 188–197.
26. Jonnalagadda S, Srinivasan R. Principal components analysis based methodology to identify differentially expressed genes in time-course microarray data. *BMC Bioinformatics*. 2008; 9(267). doi: [10.1186/1471-2105-9-267](https://doi.org/10.1186/1471-2105-9-267) PMID: [18534040](https://pubmed.ncbi.nlm.nih.gov/18534040/)
27. Huber PJ. Robust Statistics. 2nd ed. Wiley Series in Probability and Mathematical Statistics. Hoboken, NJ, USA: John Wiley and Sons; 2009. p. 200.
28. Zhao P, Yu B. On Model Selection Consistency of Lasso. *J Mach Learn Res*. 2006; 7:2541–2563.
29. Maaten LJPVD, Postma EO, Herik HJVD. Dimensionality reduction: A comparative review. Tilburg University Technical Report. 2009;1:TiCC–TR2009–005.
30. Ye J, Liu J. Sparse Methods for Biomedical Data. *SIGKDD Explor*. 2012; 14(1):4–15. doi: [10.1145/2408736.2408739](https://doi.org/10.1145/2408736.2408739) PMID: [24076585](https://pubmed.ncbi.nlm.nih.gov/24076585/)
31. Zou H, Hastie T. Regularization and variable selection via the Elastic Net. *J Roy Statist Soc Ser B*. 2005; 67(2):301–320. doi: [10.1111/j.1467-9868.2005.00503.x](https://doi.org/10.1111/j.1467-9868.2005.00503.x)
32. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*. 2010; 33(1):1–22. PMID: [20808728](https://pubmed.ncbi.nlm.nih.gov/20808728/)
33. Famakin B, Mou Y, Ruetzler C, Bembry J, Maric D, Hallenbeck J. Disruption of downstream MyD88 or TRIF Toll-like receptor signaling does not protect against cerebral ischemia. *Brain Res*; 1388:148–156. doi: [10.1016/j.brainres.2011.02.074](https://doi.org/10.1016/j.brainres.2011.02.074) PMID: [21376021](https://pubmed.ncbi.nlm.nih.gov/21376021/)
34. Brieger A, Rink L, Haase H. Differential regulation of TLR-dependent MyD88 and TRIF signaling pathways by free zinc ions. *J Immunol*. 2013; 191(4):1808–1817. doi: [10.4049/jimmunol.1301261](https://doi.org/10.4049/jimmunol.1301261) PMID: [23863901](https://pubmed.ncbi.nlm.nih.gov/23863901/)
35. Carbon S, Ireland A, Mungall CJ, Shu SQ, Marshall B, Lewis S, et al. AmiGO: online access to ontology and annotation data. *Bioinformatics*. 2008; 25(2):288–289. doi: [10.1093/bioinformatics/btn615](https://doi.org/10.1093/bioinformatics/btn615) PMID: [19033274](https://pubmed.ncbi.nlm.nih.gov/19033274/)
36. Savina A, Amigorena S. Phagocytosis and antigen presentation in dendritic cells. *Immunol Rev*. 2007; 219(1):143–156. doi: [10.1111/j.1600-065X.2007.00552.x](https://doi.org/10.1111/j.1600-065X.2007.00552.x) PMID: [17850487](https://pubmed.ncbi.nlm.nih.gov/17850487/)
37. Guermouprez P, Valladeau J, Zitvogel L, Thery C, Amigorena S. Antigen presentation and T cell stimulation by dendritic cells. *Annu Rev Immunol*. 2002; 20:621–667. doi: [10.1146/annurev.immunol.20.100301.064828](https://doi.org/10.1146/annurev.immunol.20.100301.064828) PMID: [11861614](https://pubmed.ncbi.nlm.nih.gov/11861614/)
38. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res*. 2003; 13(11):2498–2504. doi: [10.1101/gr.1239303](https://doi.org/10.1101/gr.1239303) PMID: [14597658](https://pubmed.ncbi.nlm.nih.gov/14597658/)
39. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. Gorilla: A Toll For Discovery And Visualization of Enriched GO Terms in Ranked Gene Lists. *BMC Bioinformatics*. 2009; 10(48). doi: [10.1186/1471-2105-10-48](https://doi.org/10.1186/1471-2105-10-48) PMID: [19192299](https://pubmed.ncbi.nlm.nih.gov/19192299/)
40. Sato S, Sugiyama M, Yamamoto M, Watanabe Y, Kawai T, Takeda K, et al. Toll/IL-1 Receptor Domain-Containing Adaptor Inducing IFN- $\beta$  (TRIF) Associates with TNF Receptor-Associated Factor 6 and TANK-Binding Kinase 1, and Activates Two Distinct Transcription Factors, NF- $\kappa$ B and IFN-Regulatory Factor-3, in the Toll-Like Receptor Signaling. *J Immunol*. 2003; 171(8):4304–4310. PMID: [14530355](https://pubmed.ncbi.nlm.nih.gov/14530355/)
41. Zietara N, Lyszkiewicz M, Gekara N, Puchalka J, Santos VAPMD, Hunt CR, et al. Absence of IFN-beta Impairs Antigen Presentation Capacity of Splenic Dendritic Cells via Down-Regulation of Heat Shock Protein 70. *J Immunol*. 2009; 183(2):1099–1109. doi: [10.4049/jimmunol.0803214](https://doi.org/10.4049/jimmunol.0803214) PMID: [19581626](https://pubmed.ncbi.nlm.nih.gov/19581626/)
42. Chen M, Wang YH, Wang Y, Huang L, Sandoval H, Liu YJ, et al. Dendritic Cell Apoptosis in the Maintenance of Immune Tolerance. *Science*. 2006; 331(5764):1160–1164. doi: [10.1126/science.1122545](https://doi.org/10.1126/science.1122545)
43. Ludewig B, Odermatt B, Landmann S, Hengartner H, Zinkernagel RM. Dendritic Cells Induce Autoimmune Diabetes and Maintain Disease via De Novo Formation of Local Lymphoid Tissue. *J Exp Med*. 1998; 188(8):1493–1501. doi: [10.1084/jem.188.8.1493](https://doi.org/10.1084/jem.188.8.1493) PMID: [9782126](https://pubmed.ncbi.nlm.nih.gov/9782126/)
44. Roskrow MA, Dilloo D, Suzuki N, Zhong W, Rooney CM, Brenner MK. Autoimmune disease induced by dendritic cell immunization against leukemia. *Leukemia Res*. 1999; 23(6):549–557. doi: [10.1016/S0145-2126\(99\)00045-4](https://doi.org/10.1016/S0145-2126(99)00045-4)
45. Kushwah R, Hu J. Dendritic Cell Apoptosis: Regulation of Tolerance versus Immunity. *J Immunol*. 2010; 185(2):795–802. doi: [10.4049/jimmunol.1000325](https://doi.org/10.4049/jimmunol.1000325) PMID: [20601611](https://pubmed.ncbi.nlm.nih.gov/20601611/)
46. De Trez C, Pajak B, Glaichenhaus MBN, Urbain J, Moser M, Lauvau G, et al. TLR4 and Toll-IL-1 Receptor Domain-Containing Adapter-Inducing IFN-beta, but Not MyD88, Regulate Escherichia coli-

- Induced Dendritic Cell Maturation and Apoptosis *In Vivo*. *J Immunol*. 2005; 175(2):839–846. doi: [10.4049/jimmunol.175.2.839](https://doi.org/10.4049/jimmunol.175.2.839) PMID: [16002681](https://pubmed.ncbi.nlm.nih.gov/16002681/)
47. Fuentes Marraco SA, Scott CL, Bouillet P, Ives A, Masina S, Vremec D, et al. Type I Interferon Drives Dendritic Cell Apoptosis via Multiple BH3-Only Proteins following Activation by PolyIC *In Vivo*. *PLOS ONE*. 2011; 6(6):e20189. doi: [10.1371/journal.pone.0020189](https://doi.org/10.1371/journal.pone.0020189) PMID: [21674051](https://pubmed.ncbi.nlm.nih.gov/21674051/)
  48. Lu L, Bonham CA, Chambers FG, Watkins SC, Hoffman RA, Simmons RL, et al. Induction of Nitric Oxide Synthase in Mouse Dendritic Cells by IFN- $\gamma$ , Endotoxin, and Interaction with Allogeneic T Cells: Nitric Oxide Production is Associated with Dendritic Cell Apoptosis. *J Immunol*. 1996; 157(8):3577–3586. PMID: [8871658](https://pubmed.ncbi.nlm.nih.gov/8871658/)
  49. Kaiser WJ, Offermann MK. Apoptosis Induced by the Toll-Like Receptor Adaptor TRIF Is Dependent on Its Receptor Interacting Protein Homotypic Interaction Motif. *J Immunol*. 2005; 174(8):4942–4952. doi: [10.4049/jimmunol.174.8.4942](https://doi.org/10.4049/jimmunol.174.8.4942) PMID: [15814722](https://pubmed.ncbi.nlm.nih.gov/15814722/)
  50. Rawlings JS, Rosler KM, Harrison DA. The JAK/STAT signaling pathway. *J Cell Sci*. 2004; 117:1281–1283. doi: [10.1242/jcs.00963](https://doi.org/10.1242/jcs.00963) PMID: [15020666](https://pubmed.ncbi.nlm.nih.gov/15020666/)
  51. Han J. MyD88 beyond Toll. *Nat Immunol*. 2006; 7(4):370–371. doi: [10.1038/ni0406-370](https://doi.org/10.1038/ni0406-370) PMID: [16550203](https://pubmed.ncbi.nlm.nih.gov/16550203/)
  52. Yoshimura A, Naka T, Kubo M. SOCS proteins, cytokine signaling and immune regulation. *Nat Rev Immunol*. 2007; 7(6):454–465. doi: [10.1038/nri2093](https://doi.org/10.1038/nri2093) PMID: [17525754](https://pubmed.ncbi.nlm.nih.gov/17525754/)
  53. Hayashi EA, Akira S, Nobrega A. Role of TLR in B cell development: signaling through TLR4 promotes B cell maturation and is inhibited by TLR2. *J Immunol*. 2005; 174:6639–6647. doi: [10.4049/jimmunol.174.11.6639](https://doi.org/10.4049/jimmunol.174.11.6639) PMID: [15905502](https://pubmed.ncbi.nlm.nih.gov/15905502/)
  54. Kokkonen TS, Karttunen TJ. Fas/Fas ligand-mediated apoptosis in different cell lineages and functional compartments of human lymph nodes. *J Histochem Cytochem*. 2010; 58(2):131–140. doi: [10.1369/jhc.2009.954669](https://doi.org/10.1369/jhc.2009.954669) PMID: [19826071](https://pubmed.ncbi.nlm.nih.gov/19826071/)
  55. Tang S, Chen T, Yu Z, Zhu X, Yang M, Xie B, et al. RasGRP3 limits toll-like receptor-triggered inflammatory response in macrophages by activating Rap1 small GTPase. *Nat Commun*. 2014; 5. doi: [10.1038/ncomms5657](https://doi.org/10.1038/ncomms5657)
  56. Guilbault B, Kay RJ. RasGRP1 sensitizes an immature B cell line to antigen receptor-induced apoptosis. *J Biol Chem*. 2004; 279:19523–19530. doi: [10.1074/jbc.M314273200](https://doi.org/10.1074/jbc.M314273200) PMID: [14970203](https://pubmed.ncbi.nlm.nih.gov/14970203/)
  57. Barrett DM, Gustafson KS, Wang J, Wang SZ, Ginder GD. A GATA factor mediates cell type-restricted induction of HLA-E gene transcription by gamma interferon. *Mol Cell Biol*. 2004; 24(14):6194–6204. doi: [10.1128/MCB.24.14.6194-6204.2004](https://doi.org/10.1128/MCB.24.14.6194-6204.2004) PMID: [15226423](https://pubmed.ncbi.nlm.nih.gov/15226423/)
  58. Spaggiari GM, Contini P, Dondero A, Carosio R, Puppo F, Indiveri F, et al. Soluble HLA class I induces NK cell apoptosis upon the engagement of killer-activating HLA class I receptors through FasL-Fas interaction. *Blood*. 2002; 100(12):2098–4107. doi: [10.1182/blood-2002-04-1284](https://doi.org/10.1182/blood-2002-04-1284)