Network analysis reveals cross-links of the immune pathways activated by bacteria and allergen

Colin Campbell,1,4 Juilee Thakar,2,4 and Réka Albert1,3,†

1 Department of Physics, Pennsylvania State University, University Park, Pennsylvania 16803, USA
2 Department of Pathology, Yale University School of Medicine, New Haven, Connecticut 06511, USA
3 Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16803, USA
(Received 26 April 2011; published 28 September 2011)

Many biological networks are characterized by directed edges that represent either activating (positive) or inhibiting (negative) regulation. Most graph-theoretical methods used to study biological networks either disregard this important feature, or study the role of edge sign only in the context of small subgraphs called motifs. Here, we develop path-based measures which capture, on continuous scales spanning negative and positive values, both the long- and short-range regulatory relationships among node pairs. These measures also allow the quantification of each node’s overall influence on the whole network and its susceptibility to regulation by the rest of the network. We apply the measures to a network representation of the mammalian immune response to simultaneous attack by allergen and respiratory bacteria. Although allergen and bacteria elicit different immune pathways, there is significant overlap (cross-talk) and feedback between these pathways. We identify key immune components in this cross-talk; particularly revealing the importance of natural killer cells as a key regulatory target in the cross-talk.

DOI: 10.1103/PhysRevE.84.031929 PACS number(s): 82.39.Rt, 02.10.Ox, 87.18.Mp, 87.18.Vf

I. INTRODUCTION

Extracting meaningful information from complex networks of biological interactions is a considerable challenge. A large body of work describing network measures now exists (see [1–3] for a review). For example, centrality measures such as betweenness or closeness centrality [4] are used to rank the nodes and/or edges in a network based on their importance in the global topology of the network. Though most existing measures are designed for networks wherein an edge is characterized solely by the nodes it connects, real biological networks require a more complex framework, wherein edges are additionally classified as activating (positive) or inhibiting (negative), in order to account for the endogenous and signal-induced responses of molecular and cellular species. Indeed, a variety of inhibitory mechanisms exist, including transcriptional repression (in which a transcription factor blocks the transcription of a gene) or enzyme inhibition (in which a repressor binds to the enzyme and impairs its ability to catalyze chemical reactions) [5]. Cellular and intracellular networks often contain negative regulation; examples include feedback inhibition in which the final node in a particular signaling cascade or reaction pathway inhibits the first step of the cascade, or mutual inhibition among two pathways. In immune regulatory networks bacteria, viruses, or allergens activate different types of responses which can be mutually repressive; simultaneous exposure to pathogens and allergens leads to complex positive and negative cross-regulation among immune components [6].

The impact of negative edges has been previously studied in the context of small subgraphs (typically consisting of two to four nodes) which are overexpressed in biological networks (see, for instance, [7–9]). These subgraphs, called network motifs, often have clear regulatory roles. For example, a negative feedback loop allows oscillations and a coherent feed-forward loop, i.e., two nonintersecting paths of the same sign between the same source and target node, allows for filtering of spurious activation of the target node. However, relatively little work exists which attempts to consider the role of negative edges on the scale of the entire network. Recent work has considered the effect of node knockout on the connectivity between network inputs and outputs, by implementing a network expansion procedure that takes into consideration both edge sign and conditional activation requirements [10]. However, conditional activation requirements are not always well known, and the measures proposed in [10] do not consider long-range relationships between pairs of nodes. Characterizing these relationships may offer significant insight into the structure and functioning of the network.

The most significant prior work which considers pairwise node relationships in the context of negative edges categorizes the relationship according to the existence of positive and/or negative paths between the nodes and of negative feedback loops along paths between them [11]. This classification contains six categories: activation (if there are positive paths but no negative paths), total activation (if additionally there are no nodes between them which participate in negative feedback), inhibition (if there are negative paths but no positive paths), total inhibition (if additionally there are no nodes between them which participate in negative feedback), no influence (no paths) and ambivalent (if there are positive and negative paths) [11,12]. While the first five categories allow a functional interpretation of the relationship between the two nodes, no insight is offered by the last category, ambivalent. However, there may be a significant number of networks in which many relationships are identified as ambivalent. In
order to gain insight into the short- and long-range impact of the causal relationships of the entire network, one needs a continuous approach that is able to distinguish different levels of ambivalent relationships. In this work, we develop such an approach and apply it to a network of the immune response to simultaneous attack from respiratory bacteria and allergen.

We introduce two measures of influence among nodes. The "path-based relationship" (PBR) measures consider the abundance, sign, and length of unique, non-self-intersecting paths between pairs of nodes, up to a certain length. These measures can also be used to characterize each node individually by its overall positive or negative influence on the rest of the network and vice versa. This latter function is similar to closeness centrality [4], in that it identifies nodes with small distances to many other nodes. Unlike closeness centrality, however, the PBR measures explicitly consider both the role of negative edges and the existence of multiple paths between a pair of nodes.

Due to the computational complexity associated with enumerating all paths of arbitrary length between all node pairs, the PBR measures must disregard paths longer than a certain maximum path length. This maximal value depends upon network size, density, and available computational power; in this work, the maximum length considered is 5. To consider longer-range effects, the "strength of connection" (SoC) measure considers the abundance of node-independent paths of arbitrary length between two nodes. To accommodate the increased computational complexity of analyzing longer paths, the SoC measure is based on node-independent paths; i.e., it trades detailed information about short-range node-node interactions for insight into key, long-range features of the network's topology. Both the PBR and SoC measures characterize the role of the nodes in the network in a continuous way, and thereby offer finer-grain insight into the structure of the network compared to existing measures that consider edge sign [11].

The measures are applied to an expanded version of our previously published networks of immune responses to *Bordetella* bacteria and airborne allergen [13–15]. Both bacterial infections and allergies activate a plethora of immune components, including specialized cells and signaling molecules [13–18]. However, unlike the immune response to the bacteria included in this study, which is protective (i.e., it leads to the clearance of the bacteria), allergies such as asthma involve a hypersensitive immune response against harmless substances in the environment [16,17]. Respiratory bacterial infections can influence allergic asthma and in turn allergies can affect the susceptibility of the host to respiratory infections. Based on their functional role, the immune cells can be categorized as identifiers, cells responsible for recognizing foreign invaders (such cells include dendritic cells, T0 cells and B cells) and responders, cells activated by identifier cells that control the proliferation of the invaders (such as macrophages and neutrophils). The majority of signaling molecules (such as cytokines) serve as communicators among cells [6]. An important step of the immune response is the production of proteins called antibodies that specifically bind to invaders. For this reason a general name for any substance that can lead to the activation of an immune response and the production of antibodies is "antigen" (short for antibody generator). Immune cells that aid the activation of such specific responses are called T helper cells; type I T helper cells (Th1) are predominantly activated in response to bacteria whereas type II T helper cells (Th2) are predominantly activated in response to allergen [13–18]. Th1 and Th2 responses are mutually inhibitory, but not mutually exclusive, and a large amount of cross-talk between them is present through the complex interplay between various cells and cytokines. Unlike the immune response to bacteria, which tends to be acute, controlled, and capable of generating immunological memory, the host response to the allergen is highly uncontrolled and has no memory [13–18]. The cross-talk between the immune responses can be systematically analyzed by constructing the network in which immune components are represented by nodes and signed directed edges represent the positive or negative causal relationship between them.

Our measures allow us to study the long-range interactions critical in the cross-talk between the immune pathways induced by allergen and bacteria. We combine the insights gained from our measures with the abundance of motifs in various pathways to contrast the local and global interactions. The resulting metanetwork of key node relationships identifies natural killer (NK) cells as a significant regulatory node that mediates the cross-talk between the pathways induced by bacteria and allergen.

II. METHODS

A. Path-based pairwise node relationships

We define a set of path-dependent measures which classify the relationships between nodes in regulatory networks on a continuous scale. Consider all paths (nonrepeating sequences of adjacent edges) from a specified start node to a specified end node. Each path is classified as positive if it contains 0 or an even number of negative edges and negative otherwise. We wish to cumulate these paths in a meaningful way to yield a measure of influence of the start node on the end node. In a large network, the number of paths between two nodes increases rapidly as the maximum considered path length increases. However, the significance of a path decreases as the length increases, as its regulatory influence is likely obfuscated by other behavior. Indeed, in large networks it has been shown that considering long paths is unnecessary to accurately estimate the betweenness centrality of all nodes in the network [19]. We therefore give less significance to longer paths. We only count paths whose length is less than or equal to a specified length $l_{\text{max}}$ (chosen to equal 5 in this work) since it is computationally expensive to enumerate every possible path of arbitrary length between two nodes in large complex networks. The strength of connection measure discussed in the next section considers longer paths.

To quantify the abundance of paths between two nodes without regard to their signs, we define the weighted node-node path count:

$$\omega_{ij} = \sum_{l=1}^{l_{\text{max}}} p_{ij}^+ + p_{ij}^-,$$  \hspace{1cm} (1)
Here \( p_{ij}^+ \) and \( p_{ij}^- \) are the number of positive and negative paths of length \( l \) from node \( i \) to node \( j \), respectively. We define a companion measure, the \textit{node-node path influence}, to take into account the combination of positive and negative paths:

\[
\pi_{ij} = \sum_{l=1}^{\ln} \frac{p_{ij}^+ - p_{ij}^-}{l}.
\]  

(2)

This measure will equal \( \omega_{ij} \) if node \( i \) is an activator of node \( j \) in the sense of [11], \(-\omega_{ij} \) if \( i \) is an inhibitor of \( j \), and has an intermediate value if the relationship between \( i \) and \( j \) is ambivalent. For simplicity we have chosen a linear combination with equal weights but the measure can be readily adapted to nonequal weights or nonlinear combinations.

To quantify the effect of a single node on all other nodes—or the effect of all nodes on a single node—we define the \textit{node path influence} \( \iota_i \) and \textit{node path susceptibility} \( \sigma_j \):

\[
\iota_i \equiv \sum_j \pi_{ij} \omega_{ij},
\]  

(3)

\[
\sigma_j \equiv \sum_i \pi_{ij} \omega_{ij}.
\]  

(4)

We choose the product of \( \pi_{ij} \) and \( \omega_{ij} \) as a simple way of combining the measures for a given pair of nodes: Pairs with an abundance of paths (\( \omega_{ij} \) large) which are mostly positive or negative (\(|\pi_{ij}| \) large) receive a high weight, whereas those with few paths (\( \omega_{ij} \) small) or with a mix of path types (\(|\pi_{ij}| \) small) receive low weight.

We propose an intuitive visualization of these measures and illustrate it for the simple network shown in Fig. 1(a). The network contains five nodes wherein every pair of nodes is connected by a unidirectional positive or negative edge. The fact that this network consists of many overlapping feedback and feed-forward loops is evocative of the real biological network analyzed subsequently. We characterize the role of every node pair with the measures we have defined here.

In Fig. 1(b), we characterize every pair of nodes with a circle in which the path count \( \omega_{ij} \) is indicated by the circle size and the overall path sign \( \pi_{ij}/\omega_{ij} \) is indicated by the circle color. The strongest relationships are therefore indicated by the largest circles, and the degree of their activating or inhibiting nature is immediately apparent from the color map. For instance, node C exerts less influence than the other nodes (smaller circles in row C) since its out-degree is only 1 (cf. 2 or 3 for the other nodes). Since node C has a larger in-degree than the other nodes (3 versus 2 or 1), it is influenced more than the other nodes (larger circles in column C). Few node relationships are purely positive or negative; of the 25 node pairs, four are purely positive (A \( \rightarrow \) D, C \( \rightarrow \) A, C \( \rightarrow \) D, E \( \rightarrow \) D), three are purely negative (B \( \rightarrow \) E, C \( \rightarrow \) E, E \( \rightarrow \) E), and two are overall neutral (A \( \rightarrow \) C, C \( \rightarrow \) C).

For comparison, we display the corresponding dependency matrix of Klamt \textit{et al.} in Fig. 1(d) [11]. The dependency matrix captures the node pairs with pure relationships, i.e., those which are connected by positive (negative) paths only. The dependency matrix subcategories such interactions according to whether or not the nodes on the shortest path participate in a negative feedback loop, but since every node in the sample network participates in such a loop, no distinct categorizations occur for this toy network. The remaining 18 node pairs are identified by the dependency matrix as “ambivalent” since they are connected by both positive and negative paths. The ability of the path-based relationship (PBR) measures to distinguish all 25 node pairs yields a more complete picture of the functioning of the network. For instance, node B is shown by the PBR measures to have a weak negative influence on all nodes in the network, with the exception of node D, which it weakly activates. Similarly, node C is revealed by the PBR measures to have a mixture of positive, negative, and neutral influences. The abundance of paths augments the description of purely positive or negative paths; of the four purely activating node pairs, E \( \rightarrow \) B is shown to be the most robust [largest purely green circle, Fig. 1(b)].

To visualize the node path influence and node path susceptibility for every node, we plot one against the other in Fig. 1(c). The overall role of each node is indicated by its relative position on the plot: Large horizontal positions indicate powerful positive or negative regulators, depending on direction; the vertical position indicates how strongly the node is regulated. Nodes near an axis primarily regulate or are regulated (e.g., source and sink nodes lie on the x axis and y axis, respectively), while nodes which are more evenly spaced between the axes are near the functional center of the network.

For comparison, we display the corresponding dependency matrix of Klamt \textit{et al.} in Fig. 1(d) [11]. The dependency matrix captures the node pairs with pure relationships, i.e., those which are connected by positive (negative) paths only. The dependency matrix subcategories such interactions according to whether or not the nodes on the shortest path participate in a negative feedback loop, but since every node in the sample network participates in such a loop, no distinct categorizations occur for this toy network. The remaining 18 node pairs are identified by the dependency matrix as “ambivalent” since they are connected by both positive and negative paths. The ability of the path-based relationship (PBR) measures to distinguish all 25 node pairs yields a more complete picture of the functioning of the network. For instance, node B is shown by the PBR measures to have a weak negative influence on all nodes in the network, with the exception of node D, which it weakly activates. Similarly, node C is revealed by the PBR measures to have a mixture of positive, negative, and neutral influences. The abundance of paths augments the description of purely positive or negative paths; of the four purely activating node pairs, E \( \rightarrow \) B is shown to be the most robust [largest purely green circle, Fig. 1(b)].

To visualize the node path influence and node path susceptibility for every node, we plot one against the other in Fig. 1(c). The overall role of each node is indicated by its relative position on the plot: Large horizontal positions indicate powerful positive or negative regulators, depending on direction; the vertical position indicates how strongly the node is regulated. Nodes near an axis primarily regulate or are regulated (e.g., source and sink nodes lie on the x axis and y axis, respectively), while nodes which are more evenly spaced between the axes are near the functional center of the network.
network. Node E, for instance, has a strong positive influence on the network but is influenced negatively by the network [largest positive x and negative y position in Fig. 1(c)]. In contrast, node D has a weak negative influence on the rest of the network and is positively influenced by the network.

B. Strength of connection

Since by necessity \( l_{\text{max}} \) is a relatively low value in the PBR measures, they do not consider long-range relationships. We wish to define a complementary measure which does not consider short-range relationships as comprehensively as the PBR measures but does take into account long-range relationships. To do this, we define the strength of connection (SoC) between any two nodes \( i \) and \( j \) via the following procedure:

1. Determine the shortest path and shortest path length between the nodes. If \( i = j \), consider instead the length of the shortest cycle of which the node is a part.
2. Assign the nodes on the shortest path (shortest cycle) a value equal to the shortest path (cycle) length.
3. Delete the nodes on the shortest path (cycle) aside from nodes \( i \) and \( j \) (if nodes \( i \) and \( j \) are connected, remove that edge).
4. Repeat steps (1)–(3) until nodes \( i \) and \( j \) are disconnected [nodes \( i \) and \( j \) retain the initial shortest path (cycle) length].
5. Assign any remaining nodes a value of \( \infty \).
6. Sum the inverses of the values assigned to the nodes and divide by \( N \), where \( N \) is the number of nodes in the network. If \( i = j \), normalize instead by \( \frac{N}{2} \).

The factor in step (6) is introduced to normalize the maximum possible SoC to 1: In a completely connected network, the sum of inverses when \( i \neq j \) is given by \( 2 + \frac{N-2}{2} = \frac{N+2}{2} \) (\( i \) and \( j \) share an edge and so are each assigned a value of 1, and the remaining \( N-2 \) nodes are each assigned a value of 2, since they directly connect \( i \) and \( j \)). When \( i = j \), the maximum SoC is obtained if every node forms a cycle of length 2 with the node being considered, thus the sum of inverses is \( \frac{N}{2} \).

The strength of connection as defined in the preceding procedure does not differentiate between positive and negative paths. Another possible view is to aggregate only paths of the same sign. The algorithm described here may easily be modified for this purpose by modifying step (4) to assign nodes a value of \( \infty \) if the sign of the shortest path is not the same as the original shortest path.

Since this procedure erodes the network, it does not find all paths connecting the source and target nodes; rather, it characterizes each node in the network according to the length of the shortest path that runs through it, should all shorter paths be destroyed. A high strength of connection value for a pair of nodes suggests that there is an abundance of paths running from the start node to the target node (and/or the paths are relatively short). Conversely, a low strength of connection correlates with node pairs that share few, long paths. Figure 2 shows the SoC measure applied to the toy network in Fig. 1(a). Despite the existence of an edge between every pair of nodes, the directionality yields a range of SoC values. Notably, the connection from node E to node C is among the most robust (it involves the routes E→C, E→B→C, and E→A→D→C), and the connection from node C to node E is among the least robust (it only contains the route C→A→D→E).

![FIG. 2. The strength of connection measures applied to the toy network shown in Fig. 1(a). Despite the fact that the network is strongly connected, the abundance of connections is seen to vary significantly.](031929-4.png)

C. Network assembly

We expanded our high confidence published networks of immune responses involved in allergic asthma [15] and against bacteria in the genus *Bordetellae* [13,14]. The dynamic models of these networks reproduced most of the dynamic features of allergic airway hyperresponsiveness and *Bordetellae* infections. Since new immune components have been detected to be important in these processes, we decided to expand our network by including the immune components that are known to interact with the nodes that are already present in our network [13]. This allowed us to incorporate nodes that are not yet experimentally shown to play an important role in allergic asthma or *Bordetella* infection. The expansion was performed chiefly via human inspection of the Cytokine & Cells Online Pathfinder Encyclopedia (COPE) database [20]. COPE includes indirect causal relationships such as the regulatory effects of cytokines on the production of other cytokines. Since cytokines activate the transcription of other cytokines by receptor mediated pathways in particular cell types, we only incorporated cytokine→cell→cytokine paths. The resultant network has 64 nodes and 487 edges; the nodes represent cells, cytokines, antibodies, bacteria, allergen, and phagocytosis (killing of bacteria by immune cells). To extract only the salient relationships, this network was compared to a second network, which was formed by feeding the 64 nodes into Chilibot, a PubMed abstract text mining program [21]. Each edge in the COPE-based network was accepted if the edge exists in generic form in the Chilibot-based network (e.g., if the COPE-based network contains A→B, the Chilibot-based network must contain A-B). Manual inspection of this network, coupled with a wider search of the literature, resulted in a final network with 53 nodes and 270 edges of which 32 are negative.
III. RESULTS

The network of mammalian immune responses elicited in response to allergen and bacteria is shown in Fig. 3. The majority of the network’s nodes are contained in a strongly connected component (SCC) in which every pair of nodes is connected by paths in both directions. This suggests that the two antigens induce a shared core immune response. The 12 nodes that are not part of the strongly connected component include the source nodes, allergen and bacteria, and the sink node, phagocytosis. Three nodes which are upstream from the sink node and are not part of SCC are the complexes between antibodies and bacteria, which embody the antigen-specific response elicited by bacteria. The six nodes that are immediately downstream from “bacteria” and are not part of the SCC are virulence factors of bacteria. The nodes immediately downstream of allergen are the communicator thymic stromal lymphopoietin (TSLP), and the IgE antibody-allergen complex. However, they are part of the SCC because bacteria can induce IgE and TSLP using paths of length 3.

The in- and out-degree distributions of the network are right skewed, with an average of 5.09 and maximum in-degree of 21 and out-degree of 16, respectively. The nodes with the highest in-degree include both identifiers (namely, T0 cells) and responders (namely, macrophages). Similarly, nodes having the highest out-degree include identifiers such as B cells and responders such as eosinophils. The high degree responders are corroborated by the known importance of macrophages in the inflammatory response to bacteria and the importance of eosinophils during an allergic response. Communicators typically have relatively high in- and out-degree, as expected from their role in signal propagation. Another network measure, betweenness centrality (BetCen), provides insight about the importance of nodes in the connectivity of our network. We observe a functionally diverse group of highest BetCen nodes (ranging from 0.104 to 0.155) including communicators such as histamine and transforming growth factor β (TGF-β), identifiers such as B cells, and responders such as monocytes and polymorphonuclear neutrophils (PMNs). The relative sparsity of high betweenness communicators suggests overlaps in the role of communicators, which is not surprising in this scenario. The high BetCen of B cells is due to their unique role in generating antigen-specific antibodies, which initiate a plethora of pathways. The high BetCen of PMNs and monocytes suggests that various pathways induce these two responders, which are required to finally clear the bacteria. The shortest path lengths from allergen and bacteria to the major identifier, dendritic cells, are the same, which suggests that the steps of identification are similar for both antigens. However, the shorter path length between bacteria and responders (monocytes and macrophages) (2) as compared to the path length between allergen and responders (5) suggests a faster or stronger induction of a response against bacteria.

Thus these initial analyses already suggest that responses to bacteria cross-regulate allergic responses, and that investigation of the cross-talk between the pathways activated by these two antigens is necessary in order to detect the nodes that play a critical role. However, it is not a trivial task to find the key
immune components mediating the cross-talk, given the fact that the network is dense. Moreover, one needs to take into account that 32 of the network’s 270 edges are negative. To determine the most important network components, we apply the PBR measures.

A. PBR measures reveal influential and susceptible immune components

The PBR measures cumulate short- to medium-range regulatory relationships among pairs of nodes, weighted such that shorter paths have a bigger contribution. We include paths of length 1 (edges) to 5. Figure 4 shows that the path-based node relationships are typically positive in the case of abundant paths, or nearly neutral otherwise; only a few relationships are strongly negative. Seventy-six percent of all node pairs have both positive and negative paths connecting them, and would therefore be classified as ambiguous interactions by the dependency matrix of [11]. Interestingly, of the 270 node pairs that have a direct connection (edge), 27 have a node-node influence of opposite sign (Table I). This incoherence between short- and longer-range relationships has also been encountered in incoherent feed-forward loops, which were shown to contribute to the adaptability of regulatory networks and their complex range of outputs [22]. In our network neutral and near-neutral relationships, indicating a near balance of positive and negative regulation, exist even when there are few paths between a given pair of nodes. For example, the sign of most relationships originating from the source node bacteria is nearly neutral; this indicates that bacteria, through their virulence factors, negatively regulate the immune components activated by their presence. The vast majority of node pairs that have a high path count have path influences which are clearly positive or negative: Twenty-eight percent of the node pairs have a path influence between −0.2 and 0.2, and 14% have a path count above 10% of the maximum observed value, but there are no common elements of these two groups.
Table I. The nodes in the allergy and bacteria response network which share an edge and whose normalized node-node path influence is of the opposite sign. The symbol $\text{→} (-\cdot)$ is used to indicate a positive (negative) relationship. The numerical values correspond to the coloring of circles in Fig. 4.

<table>
<thead>
<tr>
<th>Edge</th>
<th>$\omega_{ij}/\pi_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenylate cyclase toxin</td>
<td>T0 cells</td>
</tr>
<tr>
<td>B cells</td>
<td>T0 cells</td>
</tr>
<tr>
<td>Dendritic cells</td>
<td>B cells</td>
</tr>
<tr>
<td>Histamine</td>
<td>Dendritic cells</td>
</tr>
<tr>
<td>Histamine</td>
<td>Epithelial cells</td>
</tr>
<tr>
<td>Histamine</td>
<td>TREG</td>
</tr>
<tr>
<td>IL-10</td>
<td>T0 cells</td>
</tr>
<tr>
<td>IL-12</td>
<td>TH1 cells</td>
</tr>
<tr>
<td>IL-13</td>
<td>T0 cells</td>
</tr>
<tr>
<td>IL-13</td>
<td>Macrophages</td>
</tr>
<tr>
<td>IL-18</td>
<td>TH1 cells</td>
</tr>
<tr>
<td>IL-4</td>
<td>NK cells</td>
</tr>
<tr>
<td>IL-4</td>
<td>TH1 cells</td>
</tr>
<tr>
<td>IL-6</td>
<td>TREG</td>
</tr>
<tr>
<td>monocytes</td>
<td>IL-17</td>
</tr>
<tr>
<td>TGF-β</td>
<td>Dendritic cells</td>
</tr>
<tr>
<td>TGF-β</td>
<td>Epithelial cells</td>
</tr>
<tr>
<td>TGF-β</td>
<td>IGG</td>
</tr>
<tr>
<td>TGF-β</td>
<td>Macrophages</td>
</tr>
<tr>
<td>TGF-β</td>
<td>NK cells</td>
</tr>
<tr>
<td>TGF-β</td>
<td>PMNs</td>
</tr>
<tr>
<td>TREG</td>
<td>TGF-β</td>
</tr>
<tr>
<td>TREG</td>
<td>TH2 cells</td>
</tr>
<tr>
<td>TSLP</td>
<td>TH2 cells</td>
</tr>
<tr>
<td>TTSS</td>
<td>IL-10</td>
</tr>
<tr>
<td>TTSS</td>
<td>TH1 cells</td>
</tr>
<tr>
<td>TTSS</td>
<td>TH17</td>
</tr>
</tbody>
</table>

Node pairs with a negative relationship (circles with a white dot in Fig. 4) tend to have few paths; this suggests less redundancy in inhibitory regulation. Many of the immune components directly or indirectly inhibit the identifier nodes’ Th2 cells and Th17 cells, whereas most of the paths ending in Th1 cells are close to neutral. This indicates that Th2 responses are inhibited more than Th1 responses, suggesting that the immune response against bacteria (which predominantly involves Th1 responses) contains less negative regulation. The two prominent inhibitors of the network are the bacterial virulence factors’ type 3 secretion system (TTSS) and pertussis toxin. Interestingly these nodes have a negative influence on histamine and mast cells, which are immune components implicated in allergies [23].

Figure 5 depicts a scatter plot of the PBR-based node influences and susceptibilities. The node path influence and node path susceptibility are correlated with betweenness centrality in a positive, but relatively weak way (Pearson correlation coefficient of 0.57 and 0.46, respectively), which suggests that taking into account edge signs significantly modulates network connectivity. Most nodes are near the origin in quadrant 1, meaning that they regulate other nodes and are regulated by other nodes weakly and in a positive way. In contrast, Fig. 5 identifies two nodes that have significant negative influence, namely the bacterial TTSS and pertussis toxin, and two nodes that have significant negative susceptibility, namely Th2 cells and Th17 cells. The Th2 cells are critical in the generation of the observed allergic airway responses. The nodes having the highest number of incoming or outgoing edges are involved mostly in activating relationships. While an identifier node (T0 cells) and a responder node (macrophages) have the highest in-degrees and two of the three highest susceptibilities, mast cells (a responder node) have the highest susceptibility. Macrophages are usually implicated in bacterial infections whereas mast cells are the cells that release histamine, leading to allergic airway responses. NK cells also have high numbers of incoming activating paths from many different types of immune components, leading to a high susceptibility. Some other responders such as PMNs and monocytes have an abundance of both incoming and outgoing paths, which might indicate their central role in the control of bacteria and allergies; these cells stand out in Fig. 5 as the nodes having both strong susceptibility and influence. Two identifiers, B cells and dendritic cells, and two responders, eosinophils and epithelial cells, are recognized as strong positive regulators (Fig. 5). While B cells and the various types of T cells all tend to positively regulate the nodes with which they are connected (Fig. 4), the node B cells have many more such paths, and therefore have a greater node path influence (Fig. 5). This result is in agreement with the known behavior of these immune components: While T cells induce effector mechanisms, B cells are actively involved in the mechanism itself, and therefore have a stronger relation with their downstream immune components [6,13]. The four most weakly influencing nodes are those outside of the strongly connected component, upstream from phagocytosis.

Communicators tend to have a few to a moderate number of medium-range relationships and have low to intermediate positive influence and susceptibility values. The most influential communicators are interferon $\gamma$ (IFN-$\gamma$), interleukin 4 (IL-4), IL-12, IL-8, and TGF-$\beta$. IFN-$\gamma$ is connected through many medium-range interactions and one of the cell types
positively influencing IFN-γ is the responder eosinophil. It is well established that eosinophils play a critical role in allergy responses and IFN-γ plays an important role in bacterial infections [13,15]. Thus, our results suggest that an allergic response might assist bacterial clearance through its activation of the communicator IFN-γ. The intermediate susceptibility and influence of IL-12 and IL-4 is expected due to their roles in Th1 and Th2 responses against bacteria and allergen, respectively. However, the intermediate positive susceptibility and influence of TGF-β is in interesting contrast with the fact that TGF-β inhibits most of its direct targets and suggest its diverse role in regulating various immune components. Another communicator node with an intermediate susceptibility and influence is IL-8. Relatively little is known about IL-8 except for its implication in both processes activated by allergen (i.e., the release of histamine) and bacteria (i.e., the activation of macrophages). Our results corroborate the role of IL-8 in both bacterial and allergic responses since responders important in both pathogeneses positively regulate IL-8. Thus this communicator also might play a role in the cross-talk.

B. Strength of connection reveals differences in the signals transduced by allergen and bacteria

The strength of connection (SoC) indicates the abundance of paths between two nodes, without the need to disregard paths above a certain length. Since the immune network is densely connected, longer-range relationships may have a strong influence on the system dynamics. Comparing the weighted path count between two nodes (which is biased toward shorter paths) with the strength of connection between them (which includes long paths) serves as a test of the effectiveness of the former measure. We find that the SoC index is positively correlated with the weighted path count (Pearson correlation coefficient = 0.696); that is, while the measures look at different aspects of the network, their general results are in agreement (Fig. 6).

We compare the SoC of the two source nodes, allergen and bacteria, to all nodes in the network (Fig. 7), to obtain insight about the signal flow from them. The results indicate the differences in the pathways stimulated by these antigens: Almost 50% of the bacteria-induced relationships have higher SoC values than the highest SoC between allergen and nodes induced by it. This suggests that bacteria induce a much more prolific response than allergen, which is good news given that the immune response to allergen has no benefits. As expected from the important role of mast cells (responders) in allergic responses [23], mast cells have the one of the highest allergen SoC values. Interestingly, the SoC value of bacteria to mast cells is even higher; however, the relationship between them is overall negative, as the TTSS and the pertussis toxin negatively regulate mast cells (Fig. 4). The inhibition of allergic responses by bacterial infections suggested by this relationship is corroborated by the strong inhibition of the Th 2 cells (the T cell types most involved in allergies) as compared to the overall neutral regulation of the Th1 cells (the T cell types mainly elicited by bacteria) (Fig. 4). For bacteria, macrophages (which are responders) have the highest SoC value, again indicating the important role of macrophages in the immune response to bacterial infections. The decreasing SoC values between bacteria and communicators (cytokines) indicate decreasing influence; that is, some cytokines (e.g. IL-2 and IL-8) are better connected than others. In contrast, the plateau seen in the allergen distribution indicates that the various cytokines are equally weakly connected; this suggests that they play a less important role in allergies as compared to the communicator histamine, which has the third highest SoC value. The nodes having the highest SoC values in bacteria (but not allergen)-induced signaling are also prolific target nodes according to the PBR measures. In general, these results are in agreement with the PBR results (Fig. 4); for instance, there is significant overlap in the powerful bacteria-induced regulators identified by the two measures.

C. Short-range relationships versus long-range relationships

To complement the picture emerging from the medium- to long-range relationships, we study short-range interactions by identifying nodes’ participation in overrepresented network motifs. We employed the motif detection tool FANMOD, which is capable of detecting motifs consisting of up to eight nodes, and can distinguish between multiple node and edge types [24]. FANMOD detects all isomorphs of all subgraphs of a specified size in a network, and then compares their abundance to randomized networks which are generated by swapping edges in the source network. Analysis with FANMOD shows that many well known motifs such as bi-fans, feed-forward loops, negative and positive biparallel segments (both with and without feedback) are overexpressed in the network (Z score > 2). We consider only three and four node motifs that are closely related to well known motifs (e.g., a motif which differs from a bi-fan only by the existence of negative feedback) in subsequent analysis and do not count sparse subgraphs such as linear chains. Figure 8 shows the coexistence pattern of node pairs for the 30 nodes that participate in more than 250 motifs. The figure indicates that interleukins exist in many motifs. Combined with the scarcity of strong longer-range relationships (Fig. 5), the results confirm the role of communicators as short-range mediators.

FIG. 6. A scatter plot of each node pair’s weighted path count (x axis) and strength of connection (y axis). The measures are in rough agreement; a large weighted node-node path count generally corresponds to a higher strength of connection value.
NETWORK ANALYSIS REVEALS CROSS-LINKS OF THE ... PHYSICAL REVIEW E 84, 031929 (2011)

of cell-to-cell interactions. The modulation of the immune system by bacterial factors such as filamentous hemagglutinin and pertussis toxin is seen by their participation in a significant number of motifs that also include immune components. Interestingly, allergen and a few signaling molecules in the allergen pathway (e.g., IL-13) participate in more motifs than bacteria; however, this is partly due to the fact that bacterial virulence factors are shown as separate nodes.

To find the nodes that might play an important role in the cross-talk between the pathways stimulated by bacteria and allergen, we consider the number of motifs in which a pair of nodes coexists versus the weighted node-node path count between the node pair. The pairs fall in one of four groups (Fig. 9): (i) those that coexist in few motifs and have a small weighted path count (e.g., epithelial cells–Th 2 cells), (ii) those that coexist in many motifs but have a small weighted path count (e.g., adenylate cyclase toxin–allergen), (iii) those with a large weighted path count that coexist in few motifs (e.g., B cells–IL-1), and (iv) those with a significant weighted node-node path count and that coexist in a significant number of motifs (e.g., NK cells–IL-18). The node pairs in the first category are weakly connected and the node pairs in the second category have predominantly short-range interactions, whereas those in the third category have predominantly long-range interactions. The node pairs in the fourth category, that have a large weighted path count (>100) and that participate in an abundant number of motifs (>60), have both short- and long-range connections. In Fig. 9 these node pairs are shown above the dashed lines that depict the thresholds which were chosen based on the distribution of the node-node relationships and motif abundances, to isolate only the most significant relationships. This separation into distinct categories is robust to moderate changes to the threshold values have little influence on the separation into distinct categories; the separation is fundamentally different from the essentially uniform distribution obtained for the example network in Fig. 1 (shown in the inset of Fig. 9).

We find 14 node pairs connected on both short and long scales. We use these node pairs to form a connected metanetwork of the cross-talk between the allergen- and bacteria-induced pathways (Fig. 10). The nodes of this metanetwork are the 13 nodes that participate in the 14 node pairs, and each node pair is connected with an edge whose sign is determined by the sign of the corresponding path influence value. Interestingly, the responder NK cells, which can secrete communicators that interact with both Th1 and Th2 cells [18,25,26], is targeted either directly or indirectly by 12 of these relationships. Other, lower in-degree terminal nodes of the metanetwork are PMNs and T0 cells. Thus our analysis identifies NK cells as critical in the cross-talk of the pathways activated by allergen and bacteria. Among the positive regulators of NK cells, eosinophils and IL-13 are implicated in allergic

FIG. 7. The sorted strength of connection values for node pairs wherein (a) bacteria or (b) allergen are the source node.
responses [15, 23]; however, our PBR and SoC analysis shows that bacteria also have a significant long-range influence on these two nodes. IL-18 is a well-documented activator of NK cells [26]. Two virulence factors of bacteria, pertussis toxin and TTSS, are present in the metanetwork and inhibit NK cells. This opposing regulation by different components suggests that bacteria inhibit NK cells until their virulence factors are neutralized, after which NK cells might play a regulatory role in the control of the overall inflammatory responses due to their ability to secrete cytokines of both Th1 and Th2 type.

IV. DISCUSSION AND CONCLUSIONS

Networks are a useful representation of complex systems; biological networks in particular serve to greatly facilitate the integration, analysis, and interpretation of biological regulatory information. The development of network measures has grown around network frameworks that are common to most fields; negative relationships, which occur primarily in biological networks, have received little attention so far. In this paper, we have proposed topological measures which classify, in a continuous way, the regulatory impact of each node on every other node in a complex regulatory network which contains both positive (activation) and negative (inhibition) edges. Measures which classify node pairs in discrete categories run the risk of classifying many dissimilar interaction pairs into a single category. As a case study, we constructed a network of the immune response to respiratory bacterial infections and allergies. The interactions of 76% of the node pairs in this network are classified as “ambivalent” by the discrete measures proposed by Klamt et al. [11]. Thus, while discrete categorization is appropriate for many networks, there exist other networks which require a more detailed investigation. The immune network serves as a challenging biological network because it is dense; revealing the cross-links between responses generated by different infections is a nontrivial task. At the same time, a study of such networks is critical in translational research. The network studied here is induced by the simultaneous presence of allergen and bacteria which are known to induce antagonistic responses against each other. Moreover, the immune response against bacteria alone is protective whereas allergen leads to a spurious induction of the immune response. Immune networks exhibit three types of inhibitory interactions: those that originate directly from...
pathogens and inhibit the activation of the immune response, those that inhibit the immune response to a specific antigen, and those that suppress the overall immune response [6]. The measures introduced here elucidate the structure and functioning of the network formed by the pathways elicited by two antigens by revealing the effect of every node on every other node.

The PBR measures we develop in this work consider the abundance of the paths between a given pair of nodes, their lengths, and their signs. Applying these measures to all node pairs allows for rapid identification of the strength and nature of nodal interactions, as well as the overall regulatory role of each node individually. Long-range relationships influence the global connectivity of the network. The strength of connection measure considers long-range interactions, and both supports the general results obtained by the PBR measures and independently illustrates the flow of signal from bacteria and allergen, revealing that the bacteria activate a stronger response whereas allergen induces a subdued response.

Our network measures reveal causal effects which have not been tested in Bordetellae infections, for instance, inhibition of the communicator TGF-β by the virulence factors of the bacteria. Moreover, though the negative influence of the TTSS on PMNs is known, Fig. 4 indicates several redundant paths between them which include other nodes such as eosinophils, IFN-γ, monocytes, TGF-β, and B, dendritic, epithelial, NK, T, Th1, Th2, and Th17 cells. These paths serve as testable predictions of the mediators of the observed causal effect. The analysis of motifs also yields insight into as yet undocumented short-range relationships. For example, the communicator eotaxin, previously not implicated in bacterial infections, is not only present in motifs including allergen but also motifs that include bacteria. Conversely, the identifier Th17 cells participate in motifs with allergen but not with bacteria. The role of IL-18 and IL-16 is not well documented in allergies or bacterial infections. Our motif analysis suggests that these cytokines might play an equally important role in bacterial and allergic pathogenesis.

The combination of the path-based measures developed here and of analysis of network motifs reveals a metanetwork that highlights important interactions in the cross-talk between the bacteria- and allergen-induced relationships. PBR and SoC analysis suggest that bacteria elicit stronger responses (higher SoCs) and have a mixed effect on the allergic responses. For example, bacteria and their virulence factors can lead to the suppression of mast cells and IgE; however, bacteria also induce IL-13 and eosinophils which are critical in the allergic responses. Conversely, allergens might lead to the induction of cytokines (e.g., IFN-γ) required for the removal of bacteria. However, since allergen has lower SoC values it might have a relatively weak effect on the total IFN-γ production. Our combined analysis suggests that NK cells are a major target of the metanetwork. In comparison, the betweenness centrality of NK cells is 0.03, only 22% of the maximum observed betweenness centrality, ranking it the 15th highest of the 53 nodes. Indeed, the 13 nodes in the metanetwork cover a wide range of betweenness centralities, suggesting that traditional measures are not sufficient to capture complex network cross-talk. The main role of NK cells is thought to be in killing virus-infected or tumor cells, and they were not part of the bacteria and allergy seed networks from which we started. However, NK cells became part of our extended network through their interactions with, and secretion of, several communicators. Indeed, NK cells are implicated in both bacterial infections and allergic responses [18,25,26]. Based on the their ability to secrete various cytokines that can regulate different types of T helper cells, we hypothesize that NK cells are master regulators induced by both source nodes and can modulate allergic responses and aid the clearance of bacteria.
The measures proposed here can be readily applied to other biological networks, such as signal transduction or gene regulatory networks, brain neuronal networks, or any other network that includes negative relationships [27,28]. The strictly topological nature of the measures means that no kinetic information is required for their use. For dynamically robust systems where thorough dynamical analysis is nonetheless infeasible due to a lack of quantitative experimental information, using these measures, in conjunction with other network measures, enables important functional insights which can serve as a proxy for full dynamical analysis. In cases where dynamical analysis is feasible, the application of the measures defined here can inform the dynamical analysis, for example, by indicating the node pairs among which both significant positive and negative regulation exists, whose dynamics therefore may depend on kinetic details. The fact that for the network studied here all strong relationships tend to be clearly positive or negative suggests that the network’s topology channels its dynamical behavior, a conclusion reached in other contexts as well [29–32].

ACKNOWLEDGMENTS

This work was supported by the Human Frontiers of Science Program (Grant No. RGP20/2007) and a Cancer Research Institute postdoctoral fellowship to J.T.