# Elucidating Influenza Inhibition Pathways via Network Reconstruction

ARNON MAZZA,[1] IRIT GAT-VIKS,[2,*] and RODED SHARAN[1,*]

## ABSTRACT

**Viruses evade detection by the host immune system through the suppression of antiviral pathways. These pathways are thus obscured when measuring the host response to viral infection and cannot be inferred by current network reconstruction methodology. Here we aim to close this gap by providing a novel computational framework for the inference of such inhibited pathways as well as the proteins targeted by the virus to achieve this inhibition. We demonstrate the power of our method by testing it on the response to influenza infection in humans, with and without the viral inhibitory protein NS1, revealing its direct targets and their inhibitory effects.**

**Key words:** algorithms, computational molecular biology, gene networks, graphs and networks, linear programming.

## 1. INTRODUCTION

**M**ANY VIRUSES USE THE HOST's cellular machinery to replicate themselves. On the other hand, hosts contain mechanisms for sensing the presence of viruses, for example, via viral DNA/RNA sensors. These mechanisms trigger the activation of immune signaling pathways that lead to attacking the pathogen and to warning nearby cells (Takeuchi and Akira, 2009).

In order to evade detection, some viruses subvert these defense mechanisms by suppressing essential proteins or protein–protein interactions along antiviral pathways. This pattern of manipulation is sometimes driven by a particular viral protein that inhibits antiviral pathways in the host that would otherwise be induced by the other viral components. Therefore, this framework has the following key factors: a wild-type (WT) virus; some protein P of the virus carrying out an inhibitory role; and the rest of the viral components, denoted $\Delta P$. Consequently, relevant host pathways can be divided into three categories: pathways that are induced by the presence of the $\Delta P$ components; out of these, pathways that are inhibited by the WT owing to the presence of the suppressing protein P; and finally, pathways that show up only in the WT scenario, supposedly triggered by that same protein P. This general framework is illustrated in Figure 1.

In this article, we aim to elucidate these different types of pathways, focusing on the pathways that are inhibited by the special protein P. Many viruses are known to fit to this framework (Brzózka et al., 2006;

---

[1]Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel.
[2]Department of Cell Research and Immunology, Tel Aviv University, Tel Aviv, Israel.
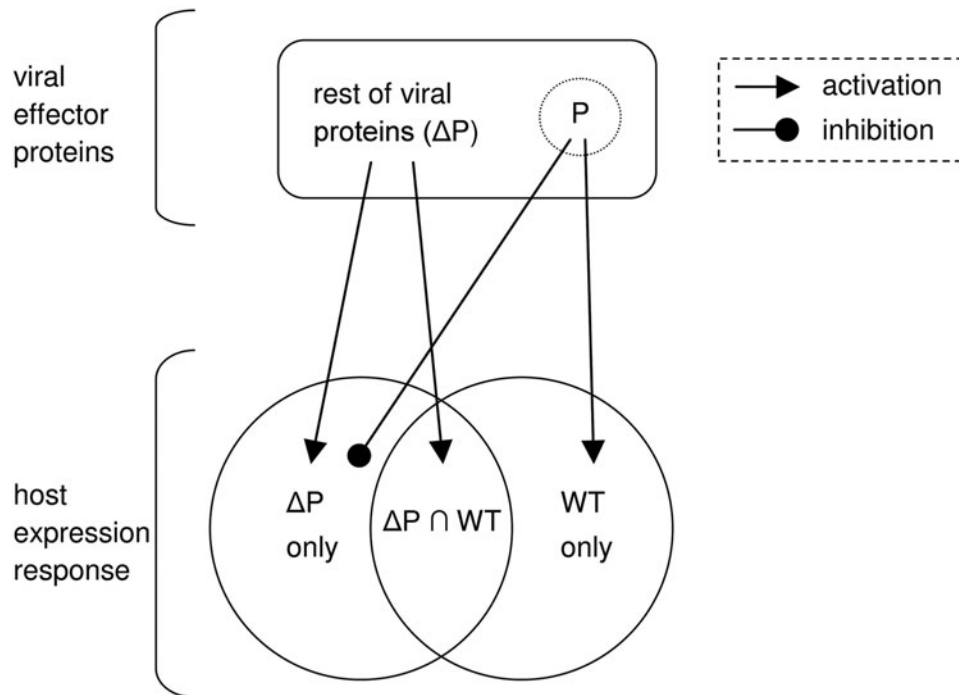*These authors contributed equally to this work.

**FIG. 1.** A schematic overview of the inhibition framework. (*Top*) Viral effector proteins, including the inhibitory protein P (e.g., NS1) and the remaining viral proteins $\Delta P$. (*Bottom*) Host transcriptional response; genes are divided into three categories: genes that are expressed in the presence of the $\Delta P$ mutant virus but not in the WT, and vice versa, and genes that are expressed in both conditions. As the WT virus includes all $\Delta P$ proteins, we would expect the differentially expressed response for the $\Delta P$ virus to be fully contained by the response to the WT virus. However, this is not the case because of the role that P plays in the inhibition of pathways that are activated by the other viral proteins.

Reid et al., 2007; Ronco et al., 1998). Here we focus on the influenza A virus, whose NS1 protein is known as an antiviral signaling inhibitor (Gack et al., 2009; Li et al., 2006; Pichlmair et al., 2006), and for which human gene expression data are available, describing the response to both the WT virus and its $\Delta NS1$ mutant.

The reconstruction of signaling pathways for a given response is based upon defining some source (aka *anchor*) proteins mediating the process, and on a genome-wide screen yielding *terminal* proteins that are affected by it. These are used to infer a compact protein–protein interaction (PPI) subnetwork connecting the anchors to the terminals (Lan et al., 2011; Yosef et al., 2011). However, the setting of the current article differs from previous ones in that the target pathway is inhibited rather than active in the studied condition, and to reveal it one should contrast the normal response to the pathogen with the response resulting from knocking out the inhibitor gene.

Our contribution in this article is twofold: (i) we provide a general computational framework for the inference of host pathways that are inhibited by a virus during an infection process; and (ii) we apply our framework to study influenza infection in humans, predicting the involvement of several human genes in the inhibition process—specifically, out of 19 inferred proteins, 7 are known to be directly inhibited by the NS1 protein and another 9 predictions are supported by evidence from other influenza proteins or from other viruses.

## 2. PRELIMINARIES

Let $G = (V, E)$ be a directed graph representing a PPI network with vertex set $V$ and edge set $E$. Let $T_2 \subset T_1 \subset V$ be two sets of terminals representing the response to the $\Delta P$ mutant ($T_1$) and the response that is common to the WT virus and its $\Delta P$ form ($T_2$). We relate to these sets as the two *conditions*. Let $a \in V \setminus T_1$ be an anchor node (either a single protein or an auxiliary node representing several proteins). We seek a

compact subtree $H = (U, F)$ of $G$ that will explain the data; that is, $H$ should span the terminals in $T_1$ and provide a parsimonious explanation for the inhibition of the terminals in $T_1 \setminus T_2$.

We assume that the virus inhibitory protein works by inhibiting specific interactions in the host through binding to one of their end points (Gack et al., 2009; Reid et al., 2007). For a vertex $v \in U$, let $H(v)$ denote the subtree rooted at $v$. If $(u, v)$ is a candidate inhibited (directed) edge in $H$, then we assume that $H(v)$ is inhibited and, hence, all the terminals within it should be in $T_1 \setminus T_2$. Thus, we seek a small subset $I \subseteq F$ of *inhibited edges* that "cover" all terminals in $T_1 \setminus T_2$ while minimizing the number of "false positive" terminals in $T_2$. We may require some minimal distance from the anchor to both endpoints of an edge for classifying that edge as inhibited—for example, if the anchor node represents all viral proteins, then a minimum distance of one reflects the reasonable assumption that only host-host interactions can be inhibited. We call the arising problem of finding a subnetwork and an associated set of inhibited edges *inhibited pathway reconstruction* (IPR). Its precise objective is defined in the next section.

## 3. THE ALGORITHM

We devised an algorithm for solving the IPR problem based on an integer linear programming (ILP) approach. The algorithm uses a flow-based technique in order to connect an anchor to a set of terminals $T_i$ by pushing $|T_i|$ units of flow out of the anchor and collecting them at the terminals. The amount of flow per edge $e \in E$ and condition $i \in \{1, 2\}$ is denoted by the variable $x_{e,i}$. The algorithm uses binary variables $q_e$ to mark the inhibited edges and uses auxiliary binary variables $r_e$ to propagate inhibition downstream, thus ensuring that no two edges on the same path from the anchor could be classified as inhibited. The variables $p_e$ assist in counting false positive terminals of $T_2$. Finally, variables of the form $z_e$ indicate for every edge $e \in E$ whether it participates in the solution subnetwork $H$.

Denote by $R \subset E$ the subset of edges touching a node that, according to prior knowledge, is able to form a physical interaction with the virus inhibitory protein. The goal is to minimize a weighted combination of: the number of edges in $H$; the number of inhibited edges from $R$; the number of inhibited edges not in $R$; and the number of false positive terminals of $T_2$. In our formulation we denote by $In(v)$ ($Out(v)$) the set of incoming (outgoing) edges of a node $v \in V$ respectively; $1_A(x)$ denotes the indicator function. The ILP formulation is given below. For clarity, we omitted the constraints on variable ranges.

$$\min \sum_{e \in E} z_e + \alpha_1 \sum_{e \in R} q_e + \alpha_2 \sum_{e \in E \setminus R} q_e + \beta \sum_{e \in E} p_e$$

s.t.:

$$\sum_{e \in Out(a)} x_{e,i} = |T_i| \qquad i = 1, 2 \tag{1}$$

$$\sum_{e \in In(v)} x_{e,i} = \sum_{e \in Out(v)} x_{e,i} + 1_{T_i}(v) \qquad \forall v \in V \setminus \{a\}, i = 1, 2 \tag{2}$$

$$x_{e,i} \le |T_i| \cdot z_e \qquad \forall e \in E, i = 1, 2 \tag{3}$$

$$\sum_{e \in In(t)} x_{e,2} = 0 \qquad \forall t \in T_1 \setminus T_2 \tag{4}$$

$$z_{e_1} + z_{e_2} \le 1 \qquad \forall v \in V, e_1 \ne e_2 \in In(v) \tag{5}$$

$$|T_2| \cdot (q_e - 1) + x_{e,2} \le p_e \le x_{e,2} \qquad \forall e \in E \tag{6}$$

$$0 \le p_e \le |T_2| \cdot q_e \qquad \forall e \in E \tag{7}$$

$$\sum_{e \in In(t)} (q_e + r_e) \ge 1 \qquad \forall t \in T_1 \setminus T_2 \tag{8}$$

$$q_e, r_e \le z_e \qquad \forall e \in E \tag{9}$$

$$q_e \le 1 - r_e \qquad \forall e \in E \tag{10}$$

$$\left( \sum_{e' \in In(v)} q_{e'} \right) + z_e - 1 \le r_e \qquad \forall v \in V, e \in Out(v) \tag{11}$$

$$\left(\sum_{e'\in In(v)} r_{e'}\right) + z_e - 1 \le r_e \qquad \forall v \in V, e \in Out(v) \tag{12}$$

$$r_e \le \sum_{e'\in In(v)} (q_{e'} + r_{e'}) \qquad \forall v \in V, e \in Out(v) \tag{13}$$

We now explain in detail the constraints of the program:

- The first two constraints ensure that each of the terminals is reached from the anchor in each of the conditions in which it is expressed: for each condition $i \in \{1, 2\}$, the anchor releases flow equal to the number of terminals observed in $T_i$; a terminal collects a single unit of flow and passes on the rest; a nonterminal passes all of its incoming flow downstream.
- Constraint (3) ensures that edges with nonzero flow take part in the solution.
- Constraint (4) requires that terminals in $T_1 \setminus T_2$ not receive flow of type $i = 2$, reflecting our biological assumption that genes expressed only in the $\Delta P$ condition do not receive an incoming signal from the anchor in the WT condition.
- Constraint (5) ensures that the resulting subnetwork forms a tree.
- The next two constraints set $p_e$ as the number of terminals falsely located under an inhibited edge $e$. That is, $p_e = x_{e,2}$ if $q_e = 1$, otherwise $p_e = 0$.
- Constraint (8) requires that all terminals in $T_1 \setminus T_2$ be inhibited, directly or indirectly.
- The last five constraints define the interrelations between $q$, $r$, and $z$. The variable $q_e$ is set to 1 if and only if the edge $e$ is inhibited directly by the virus. In this case, $r_e$ is set to 0 and the inhibition propagates to downstream neighboring edges in the solution by turning on their $r$ variable. Formally, if $q_e = 1$ then $r_e = 0$; and if $q_{u,v} = 1$ or $r_{u,v} = 1$ then for every edge $(v, w) \in E$ for which $z_{v,w} = 1$, also $r_{v,w} = 1$.

The program can be solved efficiently by noting that it is sufficient to restrict $z$ and $q$ to integer (binary) values; in this case, the other variable types ($x$, $p$, and $r$) are naturally assigned integer values in any optimal solution, as shown in the Appendix.

### 3.1. Implementation details

**Implementation platform.** We implemented the algorithm in Java, using the commercial IBM ILOG CPLEX optimizer to solve the ILP.

**ILP performance settings.** As solving the ILP is time consuming, we filtered the PPI network using a heuristic data reduction method that aims to retain those edges that are more likely to participate in the optimal solution (Mazza et al., 2013). This method selects edges that lie on a near-shortest path (in our setting up to one edge longer than a shortest path) between the anchor and any of the terminals. Further, we instructed the ILP solver to accept approximate solutions that deviate by at most 20% from the optimum and set a time limit of 3 hours for an execution. Instead of returning one arbitrary solution that fulfills the approximation ratio, we instructed the algorithm to return a solution pool of size at most 20.

**Selection of parameter values.** Our algorithm depends on three types of parameters. The first parameter is the *bias*, which reflects the difference of weights $\alpha_2 - \alpha_1$, set in the range 0–4, which we referred to as *unbiased* (0) and *biased* (1–4) experiments. For each bias we tested different combinations of the other two parameters: (i) the exact number of allowed inhibitions: either 3, 4, or 5 (added to the ILP by restricting the sum of all $q_e$-s to this value), and (ii) the false positive penalty ($\beta$): either 2, 4, or 8.

**Analyzing multiple solutions.** In order to cope with multiple solutions (either multiple approximate solutions to the same instance or solutions to different selections of parameters), all of our analyses were based on aggregate statistics rather than on a single instance. For example, to predict nodes that are likely to be located under inhibited edges, we computed per node the percent of solutions in which it is found under some inhibited edge and defined a corresponding score.

## 4. EXPERIMENTAL RESULTS

### 4.1. Data preparation

We tested our method on data from Shapira et al. (2009) concerning infection of primary human bronchial epithelial cells by the influenza virus. The data consisted of two mRNA expression profiles showing gene activity in human cells in the presence of the influenza virus: one in its wild-type form and another with the NS1 gene deleted ($\Delta NS1$ mutant). Additionally, the data contained information on physical interactions between influenza proteins and human proteins, tested using the Y2H method. We applied our algorithm to reconstruct PPI subnetworks anchored at the $\Delta NS1$ proteins through which we infer signaling pathways that are likely to be inhibited by the influenza NS1 protein.

As the background network for the application, we used the network of Yosef et al. (2011), which contains 44,738 bidirectional PPIs over 10,169 proteins. We added the $\Delta NS1$ proteins (NS2, PA, PB1, PB2, M1, M2, HA, NA, NP) along with their 124 interactions with human proteins as reported in Shapira et al. (2009). We further added manually curated interactions between the viral RNA and 16 human proteins, 5 of them having specific interactions with viral RNA (Allen et al., 2009; Diebold et al., 2004; Guillot et al., 2005; Pichlmair et al., 2006; Sabbah et al., 2009; Wu et al., 2011), and the remaining 11 interact with Poly I:C, a viral-RNA-like component (collected from the RNA-protein network of Ingenuity® Systems IPA). Finally, we connected all virus-related nodes to an artificial node that represents the $\Delta NS1$ mutant, serving as the anchor for the application. We allowed inhibitions ($e \in I$) only for edges whose both endpoints are at distances of at least two from the artificial anchor (that is, only host–host interactions).

The terminal sets for the two conditions were selected as all genes at time point $t = 8h$, whose differential expression score in the corresponding profile was above a cutoff of 0.67 (Shapira et al., 2009). We selected $t = 8$ since earlier time points had too few differentially expressed genes and later time points induced an order of magnitude more genes that might be associated with secondary responses. The first terminal set included all 62 differentially expressed genes from the $\Delta NS1$ profile, while the second set included 28 genes that were expressed in both profiles. We aimed to study how NS1 suppresses the expression of the 34 genes that belong to the first profile only. (In this study, we ignored genes in $PR8 \setminus \Delta NS1$; one possible explanation for their expression is an activation signal stemming from NS1).

We first applied the heuristic filtering, yielding 9684 directed interactions over 1666 genes. With this background, we tested all combinations of parameters, analyzed separately per bias. Every resulting solution consisted of a subnetwork and a set of inhibited edges.

### 4.2. Biological results

We used the output produced by our algorithm to provide two types of predictions. First, we sought to divide the revealed pathways into two subsets, pathways that are inhibited by NS1 (henceforth, the *inhibited set*) and pathways that are not (the *non-inhibited set*). Second, we looked for genes that are likely to play an explicit role in the NS1 inhibition process.

In the following subsections, we discuss the analyses made to achieve our predictions and describe the validation methods. In both analyses we focused on the group of experiments with bias = 4, in order to best exploit our prior knowledge on the interactions that NS1 is able to form.

*4.2.1. Extracting the inhibited pathways.* In order to distinguish between inhibited and non-inhibited pathways, we counted for each of the 232 human non-terminal genes belonging to any solution subnetwork (163 total with bias = 4 and all 9 combinations of the other two parameters), the percentage of solutions in which it appears below some inhibited edge. We assumed that the higher the percentage, the more likely it is that a gene participates in an inhibited pathway. On the other hand, we have prior knowledge that the NS1 protein has strong inhibition effects on immune system–related pathways. Therefore, to select the threshold that best divides the list into two functionally coherent groups, we applied the GSEA method (Subramanian et al., 2005) to our ranked list (using classic weights) to compute the enrichment score and leading edge subset (i.e., the high-ranking genes) with respect to the comprehensive Gene Ontology (GO) category ''cellular response to cytokine stimulus.'' This analysis yielded an enrichment score of 0.28 ($p < 0.02$ over 1000 permutations of the ranked list), breaking the list into 93 genes in the inhibited set (corresponding to a threshold of 1.2%) and 139 in the non-inhibited set.

We compared the biological functions of the inhibited versus non-inhibited gene sets using the GO annotation. We computed the enrichments of the inhibited set with respect to five antiviral response–related

| GO term | p-value | Func. set size | Int. size | Genes |
|---|---|---|---|---|
| cellular response to cytokine stimulus | 4.79E-09 | 326 | 17 | IL6ST, NFKB1, RELA, IFNAR1, IFNAR2, RIPK1, CCR3, UBC, MAPK9, TRAF2, PIN1, SUMO1, HCK, STAT3, NEDD4, JAK1, IRF3 |
| regulation of cytokine production | 2.28E-08 | 317 | 16 | IL6ST, RNF216, NFKB1, HSF1, RELA, IFNAR1, ARRB1, RIPK1, UBC, MAPK9, TRAF2, FGR, PIN1, SMAD3, KAT5, IRF3 |
| regulation of type I interferon production | 1.37E-06 | 62 | 7 | RNF216, NFKB1, RELA, RIPK1, UBC, PIN1, IRF3 |
| toll-like receptor signaling pathway | 2.07E-06 | 94 | 8 | NFKB1, RELA, MAPK1, JUN, RIPK1, UBC, MAPK9, IRF3 |
| regulation of stress-activated MAPK cascade | 0.009808 | 143 | 5 | HRAS, MAPK1, RIPK1, MAPK9, TRAF2 |

**FIG. 2.** Functional enrichment analysis of the inhibited set. For every gene ontology (GO) category (left-most column), the following information is given (columns 2–5, respectively): hyper-geometric test *p*-value, calculated based on a background of 10194 genes that comprise the human PPI network that we used; total number of genes in the GO category; size of the intersection with the 93 genes in the inhibited set; and the intersecting gene names.
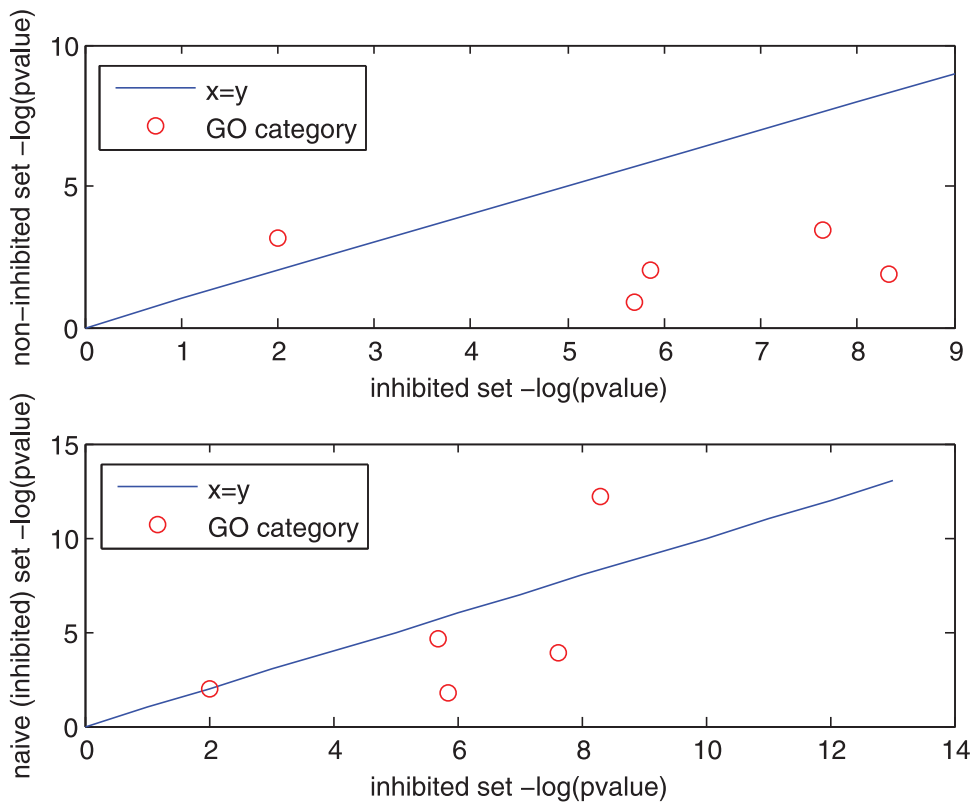


**FIG. 3.** Performance of the Inhibited Pathway Reconstruction algorithm. The plots compare the $-\log$ p-value of the hyper-geometric enrichment test between the IPR predicted inhibited pathways (x-axis) and the IPR non-inhibited pathways (top), or the inhibited pathways produced by the naïve algorithm (bottom). The circles represent GO categories. IPR, inhibited pathway reconstruction.

GO categories, such as toll-like receptor (TLR) signaling and interferon production (Fig. 2). As depicted in Figure 3a, these *p*-values were more significant than the respective *p*-values attained for the non-inhibited set, indicating that our predicted inhibited pathways indeed have a unique immune-related role, which might explain why NS1 chooses to focus on their inhibition.

Next, we compared the performance of our algorithm to that of a naïve approach, which predicts the inhibited pathways by reconstructing a minimum Steiner tree that spans the genes in $\Delta NS1 \setminus PR8$ using NS1 as the anchor. This method is inherently biased as it uses knowledge on feasible virus–host PPIs of NS1. However, it is naïve in the sense that it is not aware of the information embodied in the $\Delta NS1 \cap PR8$ condition. We applied the same heuristic filtering and collected all suggested solutions within 20% of the optimum, amounting to 44 subnetworks with a total of 93 noninput genes in their union. The intersection of these genes with the inhibited set that was inferred by our algorithm (also of size 93) contained 43 genes, showing that the predictions are quite different. Figure 3b shows that the *p*-values attained for the inhibited set produced by our algorithm compare favorably to those attained by the naïve algorithm.

*4.2.2. Inferring the inhibited genes.* To predict genes that are directly manipulated by NS1 in the inhibition process, we performed the following steps. Per algorithm execution, we gathered the inhibited edges from all suggested solutions. For each gene touching an inhibited edge (and leading to at least two terminals), we predicted it as inhibited if it showed up for at least two of the nine parameter combinations when setting the bias to a fixed value. To evaluate the robustness of these predictions with respect to the
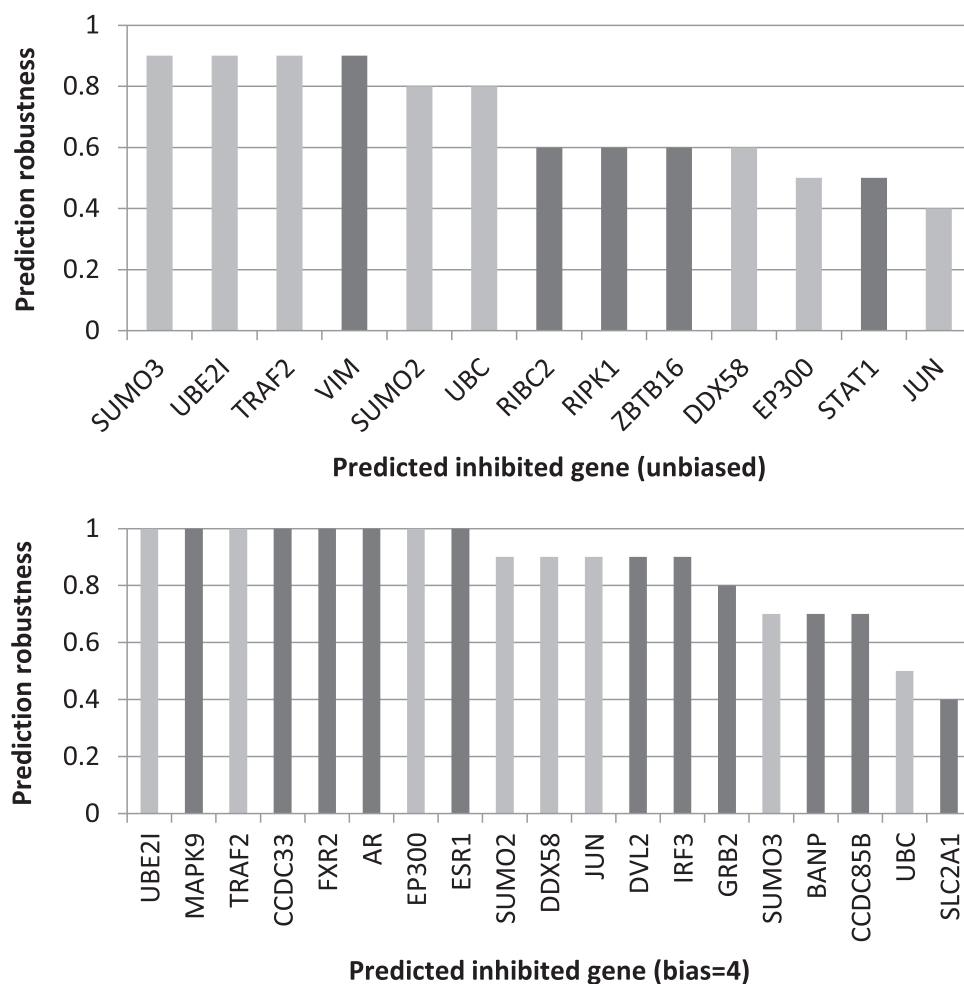


**FIG. 4.** Predicted direct inhibitions of NS1. The histograms display the prediction robustness of each gene, defined as the fraction of runs in which the gene was predicted when random 10% of the terminals were hidden. Separate histograms are shown for the unbiased and biased settings. Genes that were predicted in both settings appear in light gray.

gene expression measurements, we executed the above process 10 times, each time hiding 10% randomly chosen terminals (6 out of 62). The robustness scores of our predictions, defined as the fraction of executions in which they were raised, are summarized in Figure 4, for the unbiased (bias = 0) and biased (bias = 4) settings (we omitted three predictions whose robustness score was below 0.4). Interestingly, 8 predictions were shared between the unbiased and biased analyses out of 13 and 19 predictions in total, respectively. The predictions remained stable for a similar robustness analysis with respect to the input PPI network, performed by removing from the filtered network 5% randomly chosen interactions and adding 5% randomly chosen interactions.

We evaluated the results of the biased analysis by looking for every predicted gene, whether it is involved in any known viral interactions. Notably, 16 out of 19 genes had strong support from the literature (Fig. 5). In particular, 7 are known to interact with NS1, including UBE2I, DVL2, CCDC33, DDX58, BANP, MAPK9, and FXR2 (Mibayashi et al., 2007; Shapira et al., 2009). Another gene, TRAF2, interacts with the influenza protein PB2 (Shapira et al., 2009). Interestingly, 8 additional genes are known to interact with proteins of other viruses. For example, IRF3 interacts with the E6 protein of the human papillomavirus

| Human gene | Virus name | Virus protein | PMID | Human gene | Virus name | Virus protein | PMID |
|---|---|---|---|---|---|---|---|
| TRAF2 | H1N1 | PB2 | 20064372 | FXR2 | H1N1 | NP | 20064372 |
| TRAF2 | HHV4 | Lmp1 | 16446357 | DVL2 | H1N1 | NS1 | 20064372 |
| UBE2I | H1N1 | NS1 | 20064372 | DVL2 | H1N1 | M1 | 20064372 |
| UBE2I | H1N1 | PB1 | 20064372 | DVL2 | H1N1 | M2 | 20064372 |
| UBE2I | HHV4 | Lmp1 | 21795333 | DVL2 | H1N1 | PB1 | 20064372 |
| UBE2I | BPV1 | E1 | 10871618 | DVL2 | H1N1 | PB2 | 20064372 |
| UBE2I | Enterovirus A | 3C | 21784861 | DVL2 | HHV4 | Blrf2 | 23326445 |
| CCDC33 | H1N1 | NS1 | 20064372 | BANP | H1N1 | NS1 | 20064372 |
| CCDC33 | H1N1 | PB2 | 20064372 | BANP | H1N1 | PB1 | 20064372 |
| DDX58 | H1N1 | NS1 | 17079289 | BANP | H1N1 | PB2 | 20064372 |
| DDX58 | RVF | RNA fragment | 18446221 | EP300 | Adenovirus | E1a12s | 15741165 |
| JUN | HPV18 | DNA fragment | 10359524 | EP300 | ASFV | A238L | 18250452 |
| JUN | HTLV-1 | promoter fragment | 12386157 | EP300 | HADV5 | E1a | Ingenuity |
| MAPK9 | H1N1 | NS1 | 20064372 | EP300 | HPV16 | E6 | 10487758 |
| MAPK9 | H1N1 | PB2 | 20064372 | EP300 | HPV16 | E7 | 12970734 |
| IRF3 | HPV16 | E6 | 9649509 | SUMO2 | HIV-1 | gag-pol | 21454548 |
| IRF3 | BPV | Icp0 | 17868652 | SUMO3 | HHV4 | Ebna3c | 11739688 |
| IRF3 | Rotavirus | C-term. dom. from Nsp1 | 17765707 | SUMO3 | HIV-1 | gag-pol | 21454548 |
| FXR2 | H1N1 | NS1 | 20064372 | GRB2 | HEV | orf3 | 11518702 |
| FXR2 | H1N1 | PB1 | 20064372 | GRB2 | Mutant of AEV | Env-sea | 10066827 |
| FXR2 | H1N1 | PB2 | 20064372 | SLC2A1 | Measles virus | V | 21911578 |

**FIG. 5.** Literature support for predicted direct inhibitions of NS1. Every record shows the predicted gene, virus name, and protein name with which it interacts, as well as the Pubmed ID that reported this interaction.
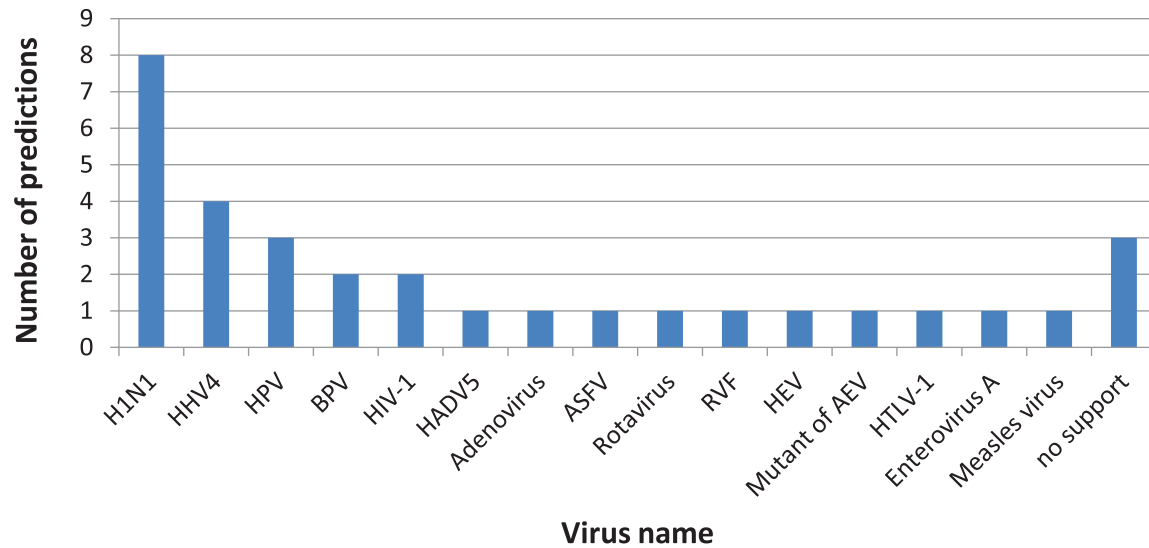
**FIG. 6.** Viral interactors of predicted inhibited genes. This histogram shows for every virus the number of predicted inhibited genes that interact with at least one protein of that virus. Predicted genes that interact with more than one type of virus are counted for each of these viruses. Predicted genes for which no support for viral interaction was found appear under the column ''no support''.

(Ronco et al., 1998), and EP300 interacts with the A238L protein of the African swine fever virus (Granja et al., 2008). As another support, we found that 7 of the 8 proteins that were predicted by both the unbiased and biased configurations have known interactions with either influenza or other viruses (Figs. 5 and 6). The genes for which we did not find support are ESR1, CCDC85B, and UBC. This might be either because such interactions are unknown or because NS1 suppresses the corresponding interactions by operating on their neighbor. To conclude, Figure 7 shows for all bias values the percentage of the predictions that are also direct NS1 interactors. The high proportion of NS1 interactors for higher bias values indicates the utility of using prior knowledge for generating more accurate and biologically relevant reconstructions.

## 5. CONCLUSIONS

In this article we described a novel network reconstruction method that is able to infer interactions and pathways that are inhibited by viral proteins. This research opens the door for various future directions. One possible extension is to relax the tree assumption, thus enabling synergy interactions and alternative pathways to genes under normal versus inhibition conditions. Another future direction is to integrate prior
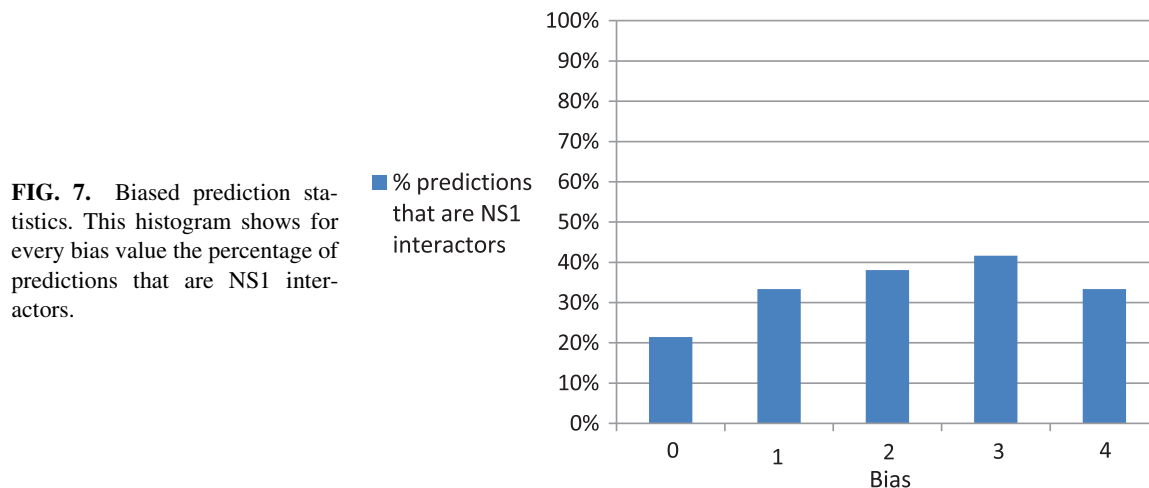
**FIG. 7.** Biased prediction statistics. This histogram shows for every bias value the percentage of predictions that are NS1 interactors.

knowledge on interaction types (activation versus inhibition) to achieve reconstructions that account for the direction of expression change. A final direction would be to explore multiple viruses simultaneously, where each virus activates different sets of differentially expressed genes while performing different inhibitions. Depending on the availability of such data, these suggested approaches may lead to new predictions and hypotheses that could be subjects for experimental validations.

# 6. APPENDIX

**Lemma 1.** *In the ILP algorithm from Section 3, assume that $z_e, q_e \in \{0, 1\}$ for every $e \in E$. Then the rest of the variable types, constrained by the ranges $r_e \in [0, 1], p_e \in [0, |T_2|], and x_{e,i} \in [0, |T_i|]$ $(i = 1, 2)$, are assigned integer values in any optimal solution.*

**Proof.** Let $v \in V$. We first prove that for every $e \in In(v)$ and $i \in \{1, 2\}$, $x_{e,i}$ is an integer. From constraint (5) and the integrality of $z$, at most one of these edges can have $z_e = 1$, and thus, from constraint (3), at most one of them can have a positive value for $x_{e,i}$. Denote by $S_i(v)$ the set of terminals reachable from $v$ in condition $i$, and let $t \in S_i(v)$. By applying the above argument for every node on the (single) path from $v$ to $t$, we conclude that $t$ can receive flow only from $v$, hence from constraint (2) we infer that $\sum_{e \in In(v)} x_{e,i} \geq |S_i(v)|$. However, the other direction holds too since the flow (for condition $i$) that $v$ sends can be collected only by terminals in $S_i(v)$. Therefore, for every $e \in In(v)$, either $x_{e,i} = 0$ or $x_{e,i} = |S_i(v)|$, as required.

The integrality of $r$ is proved by induction on the (tree) structure of an optimal subnetwork. [For an edge $e$ not in the subnetwork $r_e = 0$ from constraint (9)]. Let $v \in V$ and assume that for every $e \in In(v)$, $r_e$ is an integer. Let $f \in Out(v)$ with $z_f = 1$. Now examine the expression $\sum_{e \in In(v)} (q_e + r_e)$, which appears in the last constraint. If it evaluates to 0 we are done. Otherwise, either $\sum_{e \in In(v)} q_e > 0$ or $\sum_{e \in In(v)} r_e > 0$. Combining with the integrality of $q$, the induction hypothesis and constraint (9), either of these must equal 1. From (11) or (12) it follows that $r_f = 1$.

Finally, the integrality of $p$ follows immediately from the integrality of $q$ and $x$. ∎

# ACKNOWLEDGMENTS

# AUTHOR DISCLOSURE STATEMENT

No competing financial interests exist.

# REFERENCES

Allen, I.C., Scull, M.A., Moore, C.B., et al. 2009. The NLRP3 inflammasome mediates in vivo innate immunity to influenza A virus through recognition of viral RNA. *Immunity* 30, 556–565.

Brzózka, K., Finke, S., and Conzelmann, K.K. 2006. Inhibition of interferon signaling by rabies virus phosphoprotein P: activation-dependent binding of STAT1 and STAT2. *J Virol.* 80, 2675–2683.

Diebold, S.S., Kaisho, T., Hemmi, H., et al. 2004. Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. *Science* 303, 1529–1531.

Gack, M.U., Albrecht, R.A., Urano, T., et al. 2009. Influenza A virus NS1 targets the ubiquitin ligase TRIM25 to evade recognition by the host viral RNA sensor RIG-I. *Cell Host Microbe.* 5, 439–449.

Granja, A.G., Perkins, N.D., and Revilla, Y. 2008. A238L inhibits NF-ATc2, NF-kappa B, and c-Jun activation through a novel mechanism involving protein kinase C-theta-mediated up-regulation of the amino-terminal transactivation domain of p300. *J Immunol.* 180, 2429–2442.

Guillot, L., Le Goffic, R., Bloch, S., et al. 2005. Involvement of toll-like receptor 3 in the immune response of lung epithelial cells to double-stranded RNA and influenza A virus. *J Biol Chem.* 280, 5571–5580.

Lan, A., Smoly, I.Y., Rapaport, G., et al. 2011. ResponseNet: revealing signaling and regulatory networks linking genetic and transcriptomic screening data. *Nucleic Acids Res.* 39, W424–W429.

Li, S., Min, J.Y., Krug, R.M., et al. 2006. Binding of the influenza A virus NS1 protein to PKR mediates the inhibition of its activation by either PACT or double-stranded RNA. *Virology* 349, 13–21.

Mazza, A., Viks, I.G., Farhan, H., et al. 2013. A minimum-labeling approach for reconstructing protein networks across multiple conditions. In *Proc. of the Workshop on Algorithms in Bioinformatics (WABI)*.

Mibayashi, M., Sobrido, L.M., Loo, Y.M., et al. 2007. Inhibition of retinoic acid-inducible gene I-mediated induction of beta interferon by the NS1 protein of influenza A virus. *J Virol.* 81, 514–524.

Pichlmair, A., Schulz, O., Tan, C.P., et al. 2006. RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. *Science* 314, 997–1001.

Reid, S.P., Valmas, C., Martinez, O., et al. 2007. Ebola virus VP24 proteins inhibit the interaction of NPI-1 subfamily karyopherin alpha proteins with activated STAT1. *J Virol.* 81, 13469–13477.

Ronco, L.V., Karpova, A.Y., Vidal, M., et al. 1998. Human papillomavirus 16 E6 oncoprotein binds to interferon regulatory factor-3 and inhibits its transcriptional activity. *Genes Dev.* 12, 2061–2072.

Sabbah, A., Chang, T.H., Harnack, R., et al. 2009. Activation of innate immune antiviral responses by Nod2. *Nat Immunol.* 10, 1073–1080.

Shapira, S.D., Gat-Viks, I., Shum, B.O.V., et al. 2009. A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. *Cell* 139, 1255–1267.

Subramanian, A., Tamayo, P., Mootha, V.K., et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.*, 102, 15545–15550.

Takeuchi, O., and Akira, S. 2009. Innate immunity to virus infection. *Immunol Rev.* 227, 75–86.

The Gene Ontology Consortium. 2000. Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29.

Wu, S., Metcalf, J.P., and Wu, W. 2011. Innate immune response to influenza virus. *Curr Opin Infect Dis.* 24, 235–240.

Yosef, N., Zalckvar, E., Rubinstein, A.D., et al. 2011. ANAT: A tool for constructing and analyzing functional protein networks. *Sci. Signal.* 4, pl 1.

Address correspondence to:
*Prof. Roded Sharan*
*Blavatnik School of Computer Science*
*Tel Aviv University*
*Tel Aviv 69978*
*Israel*

*E-mail:* roded@post.tau.ac.il