Identifying proteins controlling key disease signaling pathways

Anthony Gitter$^1*$ and Ziv Bar-Joseph$^1$

$^1$School of Computer Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA

Supplementary Information

Algorithm parallelization and precomputation

SDREM was originally written as a single-threaded application, but we extended it to run on a cluster to better handle the large human datasets. In order to assess the significance of TF activity scores, SDREM generates a distribution of random TF activity scores by analyzing the gene expression data many times (typically 10 or more) using randomized TF binding interactions. Isolating these SDREM calls and executing them in parallel (approximately) reduces runtime of the IOHMM component of SDREM by the factor $\min(n, r)$, where $n$ is the number of cores available and $r$ is the number of randomizations. In the network orientation phase, we parallelized the depth first search using a synchronized priority queue to track the highest confidence paths found across all parallel threads. The source-target path enumeration tasks are divided based on the source node such that the depth first searches from each source are allocated over the available cores.

In addition, a significant speedup to the network orientation component of SDREM can be obtained by precomputing and writing all possible source-target paths to disk. In each iteration of the original version of SDREM, paths were enumerated many times because the TF connectivity was assessed by orienting a network that includes random targets, which changes the set of paths. However, it is reasonable to limit the set of potential random targets to be only TFs or even only those TFs that are present in the TF binding dataset (i.e. those TFs that could be identified as active regulators during the gene expression analysis). We now search for all paths from a source to any TF, write these paths to file, and read the appropriate stored paths for each new set of putative active targets and random TFs. Enumerating paths once instead of many times at each iteration offers immense savings computationally.

To test the impact of the approximation in which only the top $m$ paths are stored, we used the H1N1 data but only considered a high-confidence subset of the source proteins that interact with multiple viral proteins and a small set of putative TF targets so that it was possible to enumerate all paths repeatedly in a reasonable amount of time. After enumerating all $\sim 3$ million paths, we oriented the network 25 times and calculated node scores (the fraction of high-confidence paths that pass through a node) and the total path weight of the top 1000 paths for each orientation. We similarly calculated node scores and cumulative top path weight for 100 orientations in which only the 100000 or 200000 highest-confidence paths were enumerated.

$^*$Present address: Microsoft Research, 1 Memorial Drive, Cambridge, MA 02142, USA
Figure S1 shows that the actual node scores, which are used to identify which proteins participate in the signaling pathways, are highly comparable to the approximate node scores. In addition, increasing \( m \) from 100000 to 200000 does little to improve the approximation. The correlation between the actual node scores and the approximate node scores is greater than 0.999 in both cases. Similar results were obtained when using the top 100, 10000, or 50000 paths to calculate the node score instead of the top 1000.

Figure S2 shows that the top-ranked paths obtained when enumerating only \( m \) paths are not identical to those recovered when enumerating all paths. The sums of the path weights are similar, however, indicating that the sets of paths are of similar confidence. Interestingly, enumerating fewer paths results in top-ranked paths with greater cumulative weight. The low-confidence paths (that are not enumerated) no longer affect the orientation, which means that there are fewer conflicts preventing the high-confidence paths from being satisfied. This effect becomes more pronounced as a larger number of top-ranked paths are considered (e.g. 10000 and 50000), suggesting that it is preferable to consider only the top 1000 paths when limiting the number of paths that are enumerated.

Data details

We downloaded protein interaction data from BioGRID (version 3.1.74) (Stark et al., 2006) and post-translational modifications (PTMs) as well as PPI from release 9 of the Human Protein Reference Database (HPRD) (Mishra et al., 2006). Unlike the undirected BioGRID data, the PTM data provides directionality, which helps further constrain our human signaling network models. PPI and PTM weights were calculated based on the experimental methodology and number of independent detections, as in (Gitter et al., 2011). For edge \( e_{A,B} \) between proteins \( A \) and \( B \)

\[
w(e_{A,B}) = 1 - \prod_{i \in I_{A,B}} (1 - c(i))
\]

where \( w(e_{A,B}) \) is the weight of the edge, \( i \) is a member of the set \( I_{A,B} \) (all of the distinct instances of that interaction in the PPI or PTM data based on experiment type and PMID), and \( c(i) \) is the confidence in the class of experiments to which \( i \) belongs. The values of \( c(i) \) for the BioGRID interactions are taken from (Gitter et al., 2011) and reproduced in Table S1. HPRD included the more generic types of interaction evidence ‘in vivo’ and ‘in vitro’, both of which were given a confidence of 0.6. TF-gene binding predictions from (Ernst et al., 2010) were processed as described in (Schulz et al., 2012). The top 100 threshold was used in the interaction network and top 1000 threshold was used when analyzing the temporal expression data. The TF binding predictions are general predictions (not cell type specific) because the H1N1 data was aggregated from multiple cell types. Therefore we assigned them a low confidence of 0.3 in the interaction network. However, for other SDREM applications it is possible to generate condition-specific TF binding predictions using methods for integrating high-throughput data such as DNase I hypersensitivity (Neph et al., 2012).

In total, the interaction network contained 51799 PPI, 2612 PTM, and 59578 TF-gene interactions.

We downloaded the H1N1 expression data (Shapira et al., 2009) from GEO (GSE19392) (Barrett et al., 2011) using the ‘HBEs infected with PR8 post trypsin’ samples as the treatments and the ‘HBEs treated with media alone’ samples as the controls taking the average value of the two replicates at each time point before calculating fold change. For H1N1 we used the time points at 2 hours and beyond (six of the ten time points) because few transcriptional changes were observed earlier. Similarly, we downloaded the H5N1 expression data (Li et al., 2011) from GEO (GSE28166)
and took the median of the three replicates per time point before calculating fold change. For H5N1 we discarded the 0 hour time point because SDREM assumes that there is no differential expression at time point 0. The temporal expression data for H1N1 and H5N1 were filtered to include the approximately 3000 most differentially expressed genes using a log \(_2\) fold change threshold of 0.5 for H1N1 and 2.5 for H5N1.

Node priors were derived from five screens for H1N1 (Brass et al., 2009; Shapira et al., 2009; Karlas et al., 2010; König et al., 2010; Bortz et al., 2011) and one targeted screen for H5N1 (Bortz et al., 2011) that contained only 32 hits. The H1N1 host-pathogen PPI were collected from the VirHostNet database (Navratil et al., 2009) and the literature (Shapira et al., 2009; Tafforeau et al., 2011). Likewise, H5N1 sources were compiled from VirHostNet and the literature (Liu et al., 2009; Huang et al., 2009; Wang et al., 2009; Chen et al., 2010; Lee et al., 2010; Sharma et al., 2011; Tafforeau et al., 2011). In addition, we included TLR3, TLR7, TLR8, RIG-I, and NLRP3 (Koyama et al., 2007; Wang et al., 2008; Ichinohe, 2010) — proteins that either detect influenza viral RNA or influenza infection via other means — as sources for H1N1 and H5N1. The H1N1 RNAi screens affirm our assertion that screen hits are not a suitable choice for the signaling pathway source nodes because they do not capture many of the most upstream proteins involved. Of the 204 sources, only 42 (21%) are screen hits.

To test the gene prioritization algorithms’ sensitivity to the method used to identify differentially expressed genes, we also used a simple fold change heuristic instead of EDGE (Leek et al., 2006) to select the input genes and weights. For Endeavour’s input we used the sources and all genes differentially expressed at least twofold at one or more time points. For Pinta we set all genes’ weights to be the maximum magnitude of the log \(_2\) fold change over all time points. Sources were given a weight of 1, as previously recommended (Börnigen et al., 2012), if they did not already have a greater weight due to their differential expression. Note that this heuristic is less robust than EDGE’s significance analysis, which is specifically designed to account for temporal structure and dependencies (Bar-Joseph et al., 2012). Consequently, both gene prioritization algorithms perform worse using the fold change heuristic (Table S5).

**SDREM model visualization**

The regulatory pathways in Figure 1 only show a TF annotation the first time that TF is active on the path. SDREM annotates a TF on the upper path out of a split if the majority of the genes bound by the TF that pass through the split follow the upper path (likewise for the lower path). The signaling pathways were visualized with Cytoscape (Shannon et al., 2003). Although a node score threshold of 0.01 was used to generate the SDREM model, meaning that 10 of the top 1000 paths must pass through a node for it to be considered important, we relaxed this threshold to 0.005 for the model visualization and analysis. This allowed us to examine more internal nodes (for the same reason an even lower node score threshold of 0.001 was used when predicting RNAi screen hits and genetic interactions). Not all of the nodes in Table S3 appear in Figure 1 because we only draw the main connected component of the network. That is, the one TF and 103 source proteins that do not directly interact with any other proteins in Table S3 are omitted. These omitted proteins can still contribute to the response via pathways through other less important nodes that are not members of the top-ranked paths.
Running SDREM with limited data

Although we used SDREM to study the human immune response to H1N1 infection for which rich input data is available, SDREM can also be readily applied in other conditions and species with sparser datasets. Both SDREM and the original version of DREM have been applied in many species (Schulz et al., 2012; Gitter et al., 2013). Our interpretation of these studies is that the expression dataset should contain at least four time points, including the baseline time point 0, in order to run SDREM. Furthermore, we recommend SDREM even when node priors are only available for a few genes or in species where the PPI network is less complete than in human.

To demonstrate that SDREM still recovers accurate H1N1 response models with less input data, we tested it using a smaller PPI network. The restricted PPI network consisted of interactions from version 2.0.17 of BioGRID (Stark et al., 2006), the oldest archived version available, which was released in June 2006. It contains 19577 unique PPI, about half as many as the current version of BioGRID we used for all other analysis (version 3.1.74). The network is representative of the PPI networks of other organisms that have not yet been studied as extensively as human. We did not include interactions from HPRD (Mishra et al., 2006) but did include the TF-gene binding edges as before. We independently tested SDREM using a smaller set of node priors by only placing priors on the 69 genes that are hits in multiple H1N1 RNAi screens (7% of all screen hits).

Even with this limited data, SDREM still recovers many known immune response proteins including STAT1, NFKB2, and IRF transcription factors. In both the limited PPI network and limited node prior settings, SDREM’s predictions are significantly enriched for the GO term ‘immune system development’ (Benjamini-Hochberg corrected p-values 8.44 E-4 and 4.91 E-4, respectively). Overall its predictions when using the limited PPI network or limited node priors are in good agreement with the original H1N1 SDREM model that uses all available interaction and screen data (Table S7). However, as expected, SDREM performs best when it is given more complete data as input. Table S7 shows that more of SDREM’s predictions are RNAi screen hits in the original H1N1 model than in the limited data models. Note that this comparison is biased against the SDREM model that uses fewer node priors because it observes fewer screen hits as input.

To further explore the consequences of poor PPI coverage, we assessed which types of PPI data are most helpful to SDREM. The PPI that make up the highest weight paths play a special role because they are used to form SDREM’s highest confidence source-target connections. Therefore, we examined the types of evidence used to support the PPI on the top 1000 paths versus all PPI. Individual PPI can be supported by multiple experiments so we calculated the average number of times each type of evidence is associated with an interaction (Table S8). We found that although yeast two-hybrid experiments are the most abundant type of evidence overall, they are only the fourth most abundant among the PPI on the top paths. On the other hand, smaller scale experiments are enriched in the top paths. On average, the top path PPI are supported by more than one ‘Affinity Capture-Western’ experiment each. For each type of evidence we computed the enrichment in its average prevalence among the top path PPI versus all PPI. After controlling for our confidence in each type of evidence by dividing by its confidence (Table S1), we found that ‘Affinity Capture-Luminescence’ and ‘Biochemical Activity’ experiments are overrepresented among the top path PPI and ‘Protein-peptide’ experiments are underrepresented. These insights can guide the choice of PPI data to use with SDREM if a comprehensive PPI network is not available.
### Supplementary Tables

**Table S1:** Confidence scores for reported PPI and PTM

| Experiment type                  | Confidence |
|----------------------------------|------------|
| Affinity Capture-Luminescence    | 0.5        |
| Affinity Capture-MS              | 0.5        |
| Affinity Capture-RNA             | 0.7        |
| Affinity Capture-Western         | 0.5        |
| Biochemical Activity             | 0.5        |
| Co-crystal Structure             | 0.99       |
| Co-fractionation                 | 0.7        |
| Co-purification                  | 0.7        |
| Far Western                      | 0.5        |
| FRET                             | 0.7        |
| PCA                              | 0.3        |
| Protein-peptide                  | 0.7        |
| Protein-RNA                      | 0.3        |
| Reconstituted Complex            | 0.3        |
| Two-hybrid                       | 0.3        |
| In vitro                         | 0.6        |
| In vivo                          | 0.6        |

**Table S2:** Overlap among five H1N1 influenza infection RNAi screens (Brass et al., 2009; Shapira et al., 2009; Karlas et al., 2010; König et al., 2010; Bortz et al., 2011). The vast majority of the 1009 genes are hits in only a single screen.

| n | Genes detected in n screens |
|---|-----------------------------|
| 1 | 940                         |
| 2 | 62                          |
| 3 | 6                           |
| 4 | 1                           |
| 5 | 0                           |

**Table S3:** H1N1 SDREM model members. Sources are given as input, internal proteins are on the signaling paths, and targets are active TFs. Screen hits are how many of the five RNAi screens report the corresponding gene as a hit.

| Protein | Entrez gene id | Role | Screen hits |
|---------|----------------|------|-------------|
| ABLIM1  | 3983           | Source | 0           |
| ACACA   | 31             | Source | 1           |
| ACOT9   | 23597          | Source | 0           |
| ACTB    | 60             | Source | 0           |
| AIMP2   | 7965           | Source | 0           |
| ATL1    | 51062          | Source | 0           |
| ATM     | 472            | Source | 0           |
| ATP6V1G2| 534            | Source | 0           |
| BANP    | 54971          | Source | 0           |
| BCAP29  | 55973          | Source | 0           |
| Protein    | Entrez gene id | Role | Screen hits |
|------------|----------------|------|-------------|
| BHLHE40    | 8553           | Source | 1           |
| BLZF1      | 8548           | Source | 0           |
| BRD8       | 10902          | Source | 0           |
| C10ORF35   | 219738         | Source | 0           |
| C10ORF96   | 374355         | Source | 0           |
| C14ORF166  | 51637          | Source | 0           |
| C16ORF45   | 89927          | Source | 0           |
| C1ORF94    | 84970          | Source | 0           |
| C1QA       | 712            | Source | 0           |
| C7         | 730            | Source | 0           |
| CALCOCO1   | 57658          | Source | 0           |
| CAPRIN1    | 4076           | Source | 0           |
| CBS        | 875            | Source | 0           |
| CCDC33     | 80125          | Source | 0           |
| CD74       | 972            | Source | 0           |
| CDC42      | 998            | Source | 0           |
| CDC42EP4   | 23580          | Source | 0           |
| CEP152     | 22995          | Source | 0           |
| CEP70      | 80321          | Source | 0           |
| CHD6       | 84181          | Source | 0           |
| CHMP1B     | 57132          | Source | 0           |
| CHMP6      | 79643          | Source | 0           |
| CLNS1A     | 1207           | Source | 0           |
| CMTM5      | 116173         | Source | 0           |
| COL4A3BP   | 10087          | Source | 0           |
| CREB3      | 10488          | Source | 0           |
| CRK        | 1398           | Source | 0           |
| CRKL       | 1399           | Source | 0           |
| CRYAB      | 1410           | Source | 0           |
| DAZAP2     | 9802           | Source | 0           |
| DBT        | 1629           | Source | 1           |
| DDB1       | 1642           | Source | 1           |
| DDX17      | 10521          | Source | 1           |
| DDX39B     | 7919           | Source | 0           |
| DDX5       | 1655           | Source | 1           |
| DDX58      | 23586          | Source | 0           |
| DNM2       | 1785           | Source | 0           |
| DOCK8      | 81704          | Source | 0           |
| DST        | 667            | Source | 0           |
| DVL2       | 1856           | Source | 0           |
| DVL3       | 1857           | Source | 0           |
| DYNLL2     | 140735         | Source | 0           |
| EEF1A1     | 1915           | Source | 2           |
| EEF1D      | 1936           | Source | 0           |
| EIF2AK2    | 5610           | Source | 1           |
| ELP4       | 26610          | Source | 0           |
| EWSR1      | 2130           | Source | 0           |
| EXOSC8     | 11340          | Source | 0           |
| FTH1       | 2495           | Source | 0           |
| FUS        | 2521           | Source | 1           |
| FXR2       | 9513           | Source | 0           |
| GABPB1     | 2553           | Source | 0           |
| GABPB2     | 126626         | Source | 0           |
| GLYAT      | 10249          | Source | 0           |
| GLYR1      | 84656          | Source | 0           |
| GMC1L1     | 64395          | Source | 0           |
| GNB2L1     | 10399          | Source | 0           |
| Protein   | Entrez gene id | Role | Screen hits |
|-----------|----------------|------|-------------|
| GSN       | 2934           | Source 0 | 0            |
| HNRNPA1   | 3178           | Source 1 | 1            |
| HNRNPM    | 4670           | Source 1 | 1            |
| HNRNPUL1  | 11100          | Source 0 | 0            |
| HOOK1     | 51361          | Source 0 | 0            |
| HSP90AA1  | 3320           | Source 2 | 2            |
| HSP90AB1  | 3326           | Source 1 | 1            |
| HSPA1A    | 3303           | Source 1 | 1            |
| HSPA8     | 3312           | Source 1 | 1            |
| HTATSF1   | 27336          | Source 1 | 1            |
| IGHM      | 3507           | Source 0 | 0            |
| IKZF3     | 22806          | Source 0 | 0            |
| ILF3      | 3609           | Source 1 | 1            |
| IMPDH2    | 3615           | Source 1 | 1            |
| IPO5      | 3843           | Source 1 | 1            |
| IPO9      | 55705          | Source 0 | 0            |
| ITM2B     | 9445           | Source 0 | 0            |
| KARS      | 3735           | Source 0 | 0            |
| KCNRG     | 283518         | Source 0 | 0            |
| KCTD7     | 154881         | Source 0 | 0            |
| KHDRBS1   | 10657          | Source 0 | 0            |
| KHDRBS3   | 10656          | Source 0 | 0            |
| KIAA0586  | 9786           | Source 0 | 0            |
| KIAA1143  | 57456          | Source 0 | 0            |
| KPNA1     | 3836           | Source 1 | 1            |
| KPNA2     | 3838           | Source 1 | 1            |
| KPNA3     | 3839           | Source 1 | 1            |
| KPNA4     | 3840           | Source 1 | 1            |
| KPNA5     | 3841           | Source 0 | 0            |
| KPNA6     | 23633          | Source 0 | 0            |
| LNX2      | 222484         | Source 0 | 0            |
| LRRFIP1   | 9208           | Source 0 | 0            |
| LYPLA1    | 10434          | Source 0 | 0            |
| MAGEA11   | 4110           | Source 1 | 1            |
| MAGEA2    | 4101           | Source 0 | 0            |
| MAGEA2B   | 266740         | Source 0 | 0            |
| MAGEA6    | 4105           | Source 0 | 0            |
| MAGED1    | 9500           | Source 0 | 0            |
| MAPK9     | 5601           | Source 0 | 0            |
| MARS      | 4141           | Source 0 | 0            |
| MCM2      | 4171           | Source 0 | 0            |
| MCM3      | 4172           | Source 0 | 0            |
| MCM4      | 4173           | Source 0 | 0            |
| MCM5      | 4174           | Source 0 | 0            |
| MCM7      | 4176           | Source 0 | 0            |
| MEOX2     | 4223           | Source 0 | 0            |
| MGCI0075  | 84847          | Source 0 | 0            |
| MIPOL1    | 145282         | Source 0 | 0            |
| MLH1      | 4292           | Source 0 | 0            |
| MPI       | 4351           | Source 0 | 0            |
| MRI1      | 84245          | Source 0 | 0            |
| MTAP      | 4507           | Source 0 | 0            |
| NBPF22P   | 285622         | Source 0 | 0            |
| NCAFPH2   | 29781          | Source 0 | 0            |
| NCL       | 4691           | Source 1 | 1            |
| NDUFS3    | 4722           | Source 0 | 0            |
| NLRP3     | 114548         | Source 0 | 0            |
| Protein | Entrez gene id | Role | Screen hits |
|---------|----------------|------|-------------|
| NPM1    | 4869           | Source | 1           |
| NRF1    | 4899           | Source | 0           |
| NUP214  | 8021           | Source | 1           |
| NUP54   | 53371          | Source | 0           |
| NXF1    | 10482          | Source | 3           |
| NXT1    | 29107          | Source | 0           |
| OLA1    | 29789          | Source | 0           |
| PA2G4   | 5036           | Source | 1           |
| PABPC1  | 26986          | Source | 0           |
| PARP1   | 142            | Source | 1           |
| PCBD1   | 5092           | Source | 0           |
| PIK3R1  | 5295           | Source | 0           |
| PIK3R2  | 5296           | Source | 1           |
| PLAC8   | 51316          | Source | 0           |
| PNMA1   | 9240           | Source | 0           |
| POLR2A  | 5430           | Source | 0           |
| PPP2R5C | 5527           | Source | 0           |
| PRKRA   | 8575           | Source | 0           |
| PTPMT1  | 114971         | Source | 1           |
| QTRT1   | 81890          | Source | 0           |
| RABGEF1 | 27342          | Source | 0           |
| RA1      | 8480           | Source | 0           |
| RBPMS   | 11030          | Source | 0           |
| RNF5    | 6048           | Source | 0           |
| RPL11   | 6135           | Source | 0           |
| RPL5    | 6125           | Source | 0           |
| RPL8    | 6132           | Source | 0           |
| RPL9    | 6133           | Source | 0           |
| RPLP0   | 6175           | Source | 0           |
| RPS5    | 6193           | Source | 1           |
| RPS7    | 6201           | Source | 0           |
| RPS9    | 6203           | Source | 0           |
| RUVBL2  | 10856          | Source | 0           |
| SDCBP2  | 27111          | Source | 0           |
| SECISBP2| 79048          | Source | 0           |
| SEPT1   | 1731           | Source | 0           |
| SETBP1  | 26040          | Source | 0           |
| SJAH1   | 6477           | Source | 0           |
| SLC16A9 | 220963         | Source | 0           |
| SP100   | 6672           | Source | 0           |
| SRP68   | 6730           | Source | 0           |
| SRSF3   | 6428           | Source | 0           |
| SSBP2   | 23635          | Source | 0           |
| STAU1   | 6780           | Source | 0           |
| STX5    | 6811           | Source | 1           |
| TACC1   | 6867           | Source | 0           |
| TAF6    | 6878           | Source | 0           |
| TARB2   | 6895           | Source | 0           |
| TCF12   | 6938           | Source | 0           |
| TFCP2   | 7024           | Source | 0           |
| TLR3    | 7098           | Source | 0           |
| TLR7    | 51284          | Source | 0           |
| TLR8    | 51311          | Source | 0           |
| TMEM86B | 255043         | Source | 0           |
| TRAF1   | 7185           | Source | 0           |
| TRAF2   | 7186           | Source | 0           |
| TRIM25  | 7706           | Source | 1           |
| Protein   | Entrez gene id | Role | Screen hits |
|-----------|----------------|------|-------------|
| TRIM28    | 10155          | Source | 2           |
| TRIP6     | 7205           | Source | 0           |
| TTC12     | 54970          | Source | 0           |
| TUBA1B    | 10376          | Source | 0           |
| TUBB      | 203068         | Source | 1           |
| TUBB2A    | 7280           | Source | 0           |
| TUBB2C    | 10383          | Source | 0           |
| UBE2I     | 7329           | Source | 0           |
| UROS      | 7390           | Source | 0           |
| USHB1     | 83878          | Source | 0           |
| USP10     | 9100           | Source | 1           |
| UX1       | 80146          | Source | 0           |
| VIM       | 7431           | Source | 0           |
| VPS28     | 51160          | Source | 0           |
| XPO1      | 7514           | Source | 1           |
| XRCC5     | 7520           | Source | 1           |
| XRCC6     | 2547           | Source | 1           |
| YIPF6     | 286451         | Source | 0           |
| ZBTB1     | 22890          | Source | 0           |
| ZBTB25    | 7597           | Source | 0           |
| ZMAT3     | 64393          | Source | 0           |
| ZMAT4     | 79698          | Source | 1           |
| ZNF346    | 23567          | Source | 0           |
| AKT1      | 207            | Internal | 1           |
| AR        | 367            | Internal | 0           |
| CDKN1B    | 1027           | Internal | 1           |
| CHUK      | 1147           | Internal | 1           |
| CREB1     | 1385           | Internal | 1           |
| CREBBP    | 1387           | Internal | 0           |
| CTNNB1    | 1499           | Internal | 1           |
| GRB2      | 2885           | Internal | 1           |
| GSK3A     | 2931           | Internal | 1           |
| GSK3B     | 2932           | Internal | 1           |
| HIPK2     | 28996          | Internal | 1           |
| HIST3H3   | 8290           | Internal | 1           |
| JUN       | 3725           | Internal | 2           |
| KAT2A     | 2648           | Internal | 0           |
| KAT2B     | 8850           | Internal | 0           |
| MAPK1     | 5594           | Internal | 1           |
| MDM2      | 4193           | Internal | 2           |
| NFKB1     | 4790           | Internal | 1           |
| NFKBIA    | 4792           | Internal | 1           |
| NR3C1     | 2908           | Internal | 0           |
| PCNA      | 5111           | Internal | 0           |
| PRKCA     | 5578           | Internal | 1           |
| PRKCD     | 5580           | Internal | 1           |
| RUNX1     | 861            | Internal | 2           |
| RUNX2     | 860            | Internal | 0           |
| SMAD3     | 4088           | Internal | 0           |
| SMAD7     | 4092           | Internal | 1           |
| SP1       | 6667           | Internal | 0           |
| STUB1     | 10273          | Internal | 0           |
| SUMO1     | 7341           | Internal | 1           |
| TBK1      | 29110          | Internal | 1           |
| TCF3      | 6929           | Internal | 1           |
| TGFB1     | 7046           | Internal | 1           |
| TP73      | 7161           | Internal | 0           |
### Table S4: H5N1 SDREM model members. Sources are given as input, internal proteins are on the signaling paths, and targets are active TFs. Only RNAi screen hits from the small H5N1-specific screen are reported.

| Protein | Entrez gene id | Role | Screen hits |
|---------|----------------|------|-------------|
| TRIM21  | 6737           | Internal | 1           |
| UBC     | 7316           | Internal | 1           |
| AHR     | 196            | Target   | 0           |
| AIRE    | 336            | Target   | 0           |
| BRCA1   | 672            | Target   | 0           |
| DSP     | 1832           | Target   | 1           |
| E2F1    | 1869           | Target   | 1           |
| ELK1    | 2002           | Target   | 1           |
| EP300   | 2033           | Target   | 1           |
| ESR1    | 2099           | Target   | 0           |
| FOXO1   | 2308           | Target   | 0           |
| HIF1A   | 3091           | Target   | 0           |
| HSF1    | 3297           | Target   | 0           |
| IRF2    | 3660           | Target   | 2           |
| IRF3    | 3661           | Target   | 0           |
| IRF4    | 3662           | Target   | 0           |
| IRF5    | 3663           | Target   | 0           |
| IRF6    | 3664           | Target   | 1           |
| IRF7    | 3665           | Target   | 0           |
| IRF8    | 3394           | Target   | 1           |
| MYC     | 4609           | Target   | 2           |
| MYOD1   | 4654           | Target   | 1           |
| NFATC1  | 4772           | Target   | 0           |
| NFKB2   | 4791           | Target   | 0           |
| NR2F1   | 7025           | Target   | 0           |
| PPARA   | 5465           | Target   | 1           |
| RB1     | 5925           | Target   | 0           |
| RELA    | 5970           | Target   | 0           |
| SOX9    | 6662           | Target   | 0           |
| STAT1   | 6772           | Target   | 0           |
| TFAP2A  | 7020           | Target   | 1           |
| TFAP2C  | 7022           | Target   | 1           |
| TFDP1   | 7027           | Target   | 0           |
| TP53    | 7157           | Target   | 0           |
| XBP1    | 7494           | Target   | 1           |

| Protein | Entrez gene id | Role | H5N1 RNAi screen |
|---------|----------------|------|------------------|
| ATP6V1G1| 9550           | Source | N               |
| CASP8   | 841            | Source | N               |
| COMMD1  | 150684         | Source | N               |
| CPSF4   | 10898          | Source | N               |
| DDX39B  | 7919           | Source | N               |
| DDX58   | 23586          | Source | N               |
| DNASJB1 | 3337           | Source | N               |
| EIF2AK2 | 5610           | Source | N               |
| EIF4G1  | 1981           | Source | N               |
| ERBB3   | 2065           | Source | N               |
| GLUL    | 2752           | Source | N               |
| GNB2L1  | 10399          | Source | N               |
| Protein   | Entrez gene id | Role | H5N1 RNAi screen |
|-----------|----------------|------|-----------------|
| GTF3C3    | 9330           | Source | N               |
| HNRNPF    | 3185           | Source | N               |
| HSPA8     | 3312           | Source | Y               |
| ILF3      | 3609           | Source | Y               |
| IPO5      | 3843           | Source | Y               |
| IVNS1ABP  | 10625          | Source | N               |
| KPNA1     | 3836           | Source | Y               |
| KPNA2     | 3838           | Source | Y               |
| KPNA6     | 23633          | Source | N               |
| LY6D      | 8581           | Source | N               |
| MT2A      | 4502           | Source | N               |
| MX1       | 4599           | Source | N               |
| NLRP3     | 114548         | Source | N               |
| NOMO2     | 2683820        | Source | N               |
| NQO2      | 4835           | Source | N               |
| NUP98     | 4928           | Source | N               |
| PA2G4     | 5036           | Source | Y               |
| PABPN1    | 8106           | Source | N               |
| PCBP1     | 5693           | Source | N               |
| PIGQ      | 9091           | Source | N               |
| PPIA      | 5478           | Source | N               |
| PSMA7     | 5688           | Source | N               |
| SLP1      | 6590           | Source | N               |
| SNFPC4    | 6621           | Source | N               |
| STAU1     | 6780           | Source | N               |
| TLR3      | 7098           | Source | N               |
| TLR7      | 51284          | Source | N               |
| TLR8      | 51311          | Source | N               |
| XPO1      | 7514           | Source | N               |
| AR        | 367            | Internal | N               |
| ATR       | 545            | Internal | N               |
| BAG1      | 573            | Internal | N               |
| CASP3     | 836            | Internal | N               |
| CCND1     | 595            | Internal | N               |
| CDR9      | 1025           | Internal | N               |
| CHEK2     | 11200          | Internal | N               |
| CREBBP    | 1387           | Internal | N               |
| DHX9      | 1660           | Internal | N               |
| EGFR      | 1956           | Internal | N               |
| ERBB2     | 2064           | Internal | N               |
| HDAC1     | 3065           | Internal | N               |
| HDAC3     | 8841           | Internal | N               |
| HIF1A     | 3091           | Internal | N               |
| HIST3H3   | 8290           | Internal | N               |
| HSP90AA1  | 3320           | Internal | Y               |
| HSPA1A    | 3303           | Internal | Y               |
| HSPA4     | 3308           | Internal | N               |
| ING1      | 3621           | Internal | N               |
| JAK2      | 3717           | Internal | N               |
| JUN       | 3725           | Internal | N               |
| KPNB1     | 3837           | Internal | N               |
| MAPK1     | 5594           | Internal | N               |
| MAPK3     | 5595           | Internal | N               |
| MDM2      | 4193           | Internal | N               |
| MED1      | 5469           | Internal | N               |
| NCOA1     | 8648           | Internal | N               |
| NCOA2     | 10499          | Internal | N               |
| Protein | Entrez gene id | Role | H5N1 RNAi screen |
|---------|---------------|------|------------------|
| NCOA3   | 8202          | Internal | N               |
| NCOA6   | 23054         | Internal | N               |
| NCOR2   | 9612          | Internal | N               |
| NFkBIA  | 4792          | Internal | N               |
| NPM1    | 4869          | Internal | Y               |
| NR3C1   | 2908          | Internal | N               |
| NRIP1   | 8204          | Internal | N               |
| PARP1   | 142           | Internal | Y               |
| POU2F1  | 5451          | Internal | N               |
| PPP1CA  | 5499          | Internal | N               |
| PRKDC   | 5591          | Internal | N               |
| PRMT1   | 3276          | Internal | N               |
| RAF1    | 5894          | Internal | N               |
| RELA    | 5970          | Internal | N               |
| SIN3A   | 25942         | Internal | N               |
| SMAD3   | 4088          | Internal | N               |
| SMARCA4 | 6597          | Internal | N               |
| SRC     | 6714          | Internal | N               |
| STAT1   | 6772          | Internal | N               |
| STUB1   | 10273         | Internal | N               |
| SUMO1   | 7341          | Internal | N               |
| SUMO4   | 387082        | Internal | N               |
| UBC     | 7316          | Internal | N               |
| ALX1    | 8092          | Target  | N               |
| ATF4    | 468           | Target  | N               |
| BRCA1   | 672           | Target  | N               |
| CASR    | 846           | Target  | N               |
| CREB1   | 1385          | Target  | N               |
| CREM    | 1390          | Target  | N               |
| DSP     | 1832          | Target  | N               |
| E2F1    | 1869          | Target  | N               |
| EGR1    | 1958          | Target  | N               |
| EP300   | 2033          | Target  | N               |
| ESR1    | 2099          | Target  | N               |
| GABPB1  | 2553          | Target  | N               |
| GATA1   | 2623          | Target  | N               |
| GATA2   | 2624          | Target  | N               |
| GATA3   | 2625          | Target  | N               |
| GTF2A1  | 2957          | Target  | N               |
| GTF2A2  | 2958          | Target  | N               |
| HES1    | 3280          | Target  | N               |
| HINFP   | 25988         | Target  | N               |
| HSF1    | 3297          | Target  | N               |
| IRF9    | 10379         | Target  | N               |
| NR1H3   | 10062         | Target  | N               |
| NR1H2   | 8856          | Target  | N               |
| NR2F1   | 7025          | Target  | N               |
| NR2F2   | 7026          | Target  | N               |
| NR5A1   | 2516          | Target  | N               |
| PAX2    | 5076          | Target  | N               |
| RB1     | 5925          | Target  | N               |
| REL     | 5966          | Target  | N               |
| RUNX2   | 860           | Target  | N               |
| RXRA    | 6256          | Target  | N               |
| RXRB    | 6257          | Target  | N               |
| SP1     | 6667          | Target  | N               |
| STAT3   | 6774          | Target  | N               |
| Protein | Entrez gene id | Role  | H5N1 RNAi screen |
|---------|----------------|-------|------------------|
| STAT4   | 6775           | Target| N                |
| STAT5B  | 6777           | Target| N                |
| TFAP2A  | 7020           | Target| N                |
| TP53    | 7157           | Target| N                |
| VDR     | 7421           | Target| N                |
| YY1     | 7528           | Target| N                |
Table S5: Comparison of SDREM, Endeavour, and Pinta gene rankings when using fold change to identify differentially expressed genes. In this setting SDREM’s advantage over Endeavour and Pinta is even greater than when using EDGE to select the input genes.

| Algorithm | Settings                  | Hits in top 10 | Hits in top 20 | Hits in top 50 | Hits in top 100 |
|-----------|---------------------------|----------------|----------------|----------------|-----------------|
| SDREM     | Top, Targets, Weighted    | 6              | 8              | 18             | 42              |
| Endeavour | All evidence              | 2              | 4              | 10             | 24              |
| Pinta     | Default                   | 3              | 6              | 13             | 19              |

Table S6: The top 10 predicted H5N1 genetic interactions.

| Gene A | Gene B | $\epsilon_{AB}$ | $P_{AB}^{o}$ | $P_{AB}^{e}$ | $P_{A}^{o}$ | $P_{B}^{o}$ |
|--------|--------|------------------|--------------|--------------|-------------|-------------|
| HSPA8  | PA2G4  | -0.0435          | 0.5798       | 0.6233       | 0.7647      | 0.8151      |
| HSPA8  | AR     | -0.0370          | 0.6025       | 0.6396       | 0.7647      | 0.8363      |
| HSPA8  | ILF3   | -0.0234          | 0.6655       | 0.6888       | 0.7647      | 0.9008      |
| HSPA8  | KPNA2  | -0.0199          | 0.6801       | 0.7000       | 0.7647      | 0.9154      |
| ILF3   | PA2G4  | -0.0184          | 0.7159       | 0.7342       | 0.9008      | 0.8151      |
| ILF3   | AR     | -0.0162          | 0.7371       | 0.7533       | 0.9008      | 0.8363      |
| KPNA2  | PA2G4  | -0.0156          | 0.7305       | 0.7462       | 0.9154      | 0.8151      |
| GNB2L1 | HSPA8  | -0.0148          | 0.7018       | 0.7166       | 0.9371      | 0.7647      |
| ESR1   | PA2G4  | -0.0142          | 0.7261       | 0.7403       | 0.9082      | 0.8151      |
| HSPA8  | CASP8  | -0.0142          | 0.7045       | 0.7187       | 0.7647      | 0.9398      |

Table S7: Comparison of the original H1N1 SDREM model that uses all available input data and variants that use restricted versions of the input data. The limited PPI model uses an older, smaller BioGRID PPI network. The limited node prior model only places node priors on genes appearing as hits in multiple RNAi screens. Predictions include proteins on the internal signaling paths and TFs but not the source proteins given as input, which are the same in all models.

|                                           | Original model | Limited PPI model | Limited node prior model |
|-------------------------------------------|----------------|-------------------|--------------------------|
| Predicted proteins                        | 69             | 84                | 62                       |
| Predictions in RNAi screen hits           | 55%            | 37%               | 31%                      |
| Predictions in original model             | -              | 44%               | 66%                      |
Table S8: The average number of times each type of evidence is used to support a PPI. Top path PPI are the 447 PPI that make up the 1000 paths with the highest weight in the SDREM H1N1 model. Enrichment is the ratio of evidence in the top path PPI versus all PPI. The normalized version divides by edge confidence (Table S1)

| PPI evidence                  | All PPI | Top path PPI | Top path enrichment | Normalized enrichment |
|-------------------------------|---------|--------------|---------------------|-----------------------|
| Affinity Capture-Luminescence | 0.000135| 0.00224      | 16.55               | 33.11                 |
| Biochemical Activity          | 0.0234  | 0.139        | 5.92                | 11.85                 |
| Affinity Capture-Western      | 0.288   | 1.49         | 5.15                | 10.30                 |
| Reconstituted Complex         | 0.197   | 0.579        | 2.95                | 9.82                  |
| PCA                           | 0.00189 | 0.00447      | 2.36                | 7.88                  |
| Co-fractionation              | 0.00635 | 0.0268       | 4.23                | 6.04                  |
| Co-purification               | 0.0152  | 0.0537       | 3.54                | 5.05                  |
| Far Western                   | 0.00571 | 0.0134       | 2.35                | 4.70                  |
| Co-crystal Structure          | 0.00488 | 0.0201       | 4.12                | 4.16                  |
| FRET                          | 0.00185 | 0.00447      | 2.41                | 3.45                  |
| In vitro                      | 0.248   | 0.512        | 2.06                | 3.44                  |
| In vivo                       | 0.121   | 0.148        | 1.22                | 2.04                  |
| Two-hybrid                    | 0.301   | 0.181        | 0.60                | 2.01                  |
| Affinity Capture-MS           | 0.168   | 0.136        | 0.81                | 1.62                  |
| Protein-peptide               | 0.00394 | 0.00224      | 0.57                | 0.81                  |
| Affinity Capture-RNA          | 0.000174| 0            | 0                   | 0                     |
| Protein-RNA                   | 0.000135| 0            | 0                   | 0                     |
Supplementary Figures

Figure S1: Node scores, the fraction of the top 1000 paths that pass through a particular protein, are very similar when enumerating all paths or only the top 100,000 paths. The node score obtained when using all paths is shown along the x-axis. The y-axis provides the approximate node score.
Figure S2: Histograms of the sum of the path weights for the top 1000 paths (the number of paths used to calculate node scores). The blue histogram shows the distribution of the cumulative top path weights when all paths are enumerated. The red histogram corresponds to the approximation where only 100000 paths are used. Note that only 25 runs were used to generate blue histogram versus 100 for the red histogram, accounting for the taller peaks in the red histogram.
References

Bar-Joseph, Z. et al. (2012) Studying and modelling dynamic biological processes using time-series gene expression data. Nature Reviews Genetics, 13 (8), 552–564.

Barrett, T. et al. (2011) NCBI GEO: archive for functional genomics data sets–10 years on. Nucleic Acids Research, 39 (suppl 1), D1005–D1010.

Börnigen, D. et al. (2012) An unbiased evaluation of gene prioritization tools. Bioinformatics, 28 (23), 3081–3088.

Bortz, E. et al. (2011) Host- and strain-specific regulation of influenza virus polymerase activity by interacting cellular proteins. mBio, 2 (4), e00151–11.

Brass, A.L. et al. (2009) The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. Cell, 139 (7), 1243–1254.

Chen, J. et al. (2010) Human cellular protein nucleoporin hNup98 interacts with influenza A virus NS2/nuclear export protein and overexpression of its GLFG repeat domain can inhibit virus propagation. Journal of General Virology, 91 (10), 2474–2484.

Ernst, J. et al. (2010) Integrating multiple evidence sources to predict transcription factor binding in the human genome. Genome Research, 20 (4), 526 –536.

Gitter, A. et al. (2011) Discovering pathways by orienting edges in protein interaction networks. Nucleic Acids Research, 39 (4), e22.

Gitter, A. et al. (2013) Linking the signaling cascades and dynamic regulatory networks controlling stress responses. Genome Research, 23 (2), 365–376.

Huang, S. et al. (2009) Influenza A virus matrix protein 1 interacts with hTFIIIC102-s, a short isoform of the polypeptide 3 subunit of human general transcription factor IIIC. Archives of Virology, 154 (7), 1101–1110.

Ichinohe, T. (2010) Respective roles of TLR, RIG-I and NLRP3 in influenza virus infection and immunity: impact on vaccine design. Expert Review of Vaccines, 9 (11), 1315–1324.

Karlas, A. et al. (2010) Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. Nature, 463 (7282), 818–822.

Konig, R. et al. (2010) Human host factors required for influenza virus replication. Nature, 463 (7282), 813–817.

Koyama, S. et al. (2007) Differential role of TLR- and RLR-signaling in the immune responses to influenza A virus infection and vaccination. The Journal of Immunology, 179 (7), 4711–4720.

Lee, J.H. et al. (2010) Direct interaction of cellular hnRNP-F and NS1 of influenza A virus accelerates viral replication by modulation of viral transcriptional activity and host gene expression. Virology, 397 (1), 89–99.

Leck, J.T. et al. (2006) EDGE: extraction and analysis of differential gene expression. Bioinformatics, 22 (4), 507 –508.

Li, C. et al. (2011) Host regulatory network response to infection with highly pathogenic H5N1 avian influenza virus. Journal of Virology, 85 (21), 10955–10967.

Liu, D. et al. (2009) Interspecies transmission and host restriction of avian H5N1 influenza virus. Science in China Series C: Life Sciences, 52 (5), 428–438.

Mishra, G.R. et al. (2006) Human protein reference database–2006 update. Nucleic Acids Research, 34 (suppl 1), D411–414.

Navratil, V. et al. (2009) VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus-host interaction networks. Nucleic Acids Research, 37 (suppl 1), D661–D668.

Neph, S. et al. (2012) An expansive human regulatory lexicon encoded in transcription factor footprints. Nature, 489 (7414), 83–90.

Schulz, M.H. et al. (2012) DREM 2.0: improved reconstruction of dynamic regulatory networks from time-series expression data. BMC Systems Biology, 6 (1), 104.
Shannon, P. et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, **13** (11), 2498–2504.

Shapira, S.D. et al. (2009) A physical and regulatory map of host-influenza interactions reveals pathways in H1N1 infection. *Cell*, **139** (7), 1255–1267.

Sharma, K. et al. (2011) Influenza A virus nucleoprotein exploits Hsp40 to inhibit PKR activation. *PLoS ONE*, **6** (6), e20215.

Stark, C. et al. (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Research*, **34** (suppl 1), D535–539.

Tafforeau, L. et al. (2011) Generation and comprehensive analysis of an influenza virus polymerase cellular interaction network. *Journal of Virology*, **85** (24), 13010–13018.

Wang, J.P. et al. (2008) Toll-like receptor-mediated activation of neutrophils by influenza A virus. *Blood*, **112** (5), 2028–2034.

Wang, P. et al. (2009) Nuclear factor 90 negatively regulates influenza virus replication by interacting with viral nucleoprotein. *Journal of Virology*, **83** (16), 7850–7861.