Supplementary Material

Main paper: UW Statistics Technical Report 1174 (v2)
Authors: Z Wang, Q He, B. Larget, and MA Newton
Version: October 23, 2013
Contacts: Zhishi Wang, `wangz@stat.wisc.edu`
Michael A. Newton, `newton@stat.wisc.edu`

- Section 1: evaluating violation probabilities
- Section 2: estimating false-positive and true-positive rates
- Section 3: algorithm for trimming list of active sets
- Table S1: MFA-ILP results in T2D example
- Table S2: MGSA results in T2D example
- Table S3: Fisher results in T2D example
- Table S4: MFA-MCMC sets with high posterior probability, T2D
- Table S5: MFA-ILP results in influenza RNAi example
- Table S6: MGSA results in influenza RNAi example
- Table S7: Fisher results in influenza RNAi example
- Table S8: MFA-MCMC sets with high posterior probability, RNAi
- Table S9: Summary of coverage and mis-coverage, two examples
- Figure S1: KEGG pathways identified in T2D example
- Figure S2: KEGG pathways identified in influenza RNAi example

1 An algorithm to evaluate the violation probability $P_1(V_w = 1)$

A violation of AH occurs for whole w if (and only if) $Z_w = 0$ and yet $A_p = 1$ for all $p \in w$. Immediately, we see $P_1(V_w = 1) = (1 - \pi)P\left(A_p = 1 \forall p \in w | Z_w = 0\right)$. The second factor is difficult to compute by full enumeration since all configurations of overlapping sets must be considered, and this is exponential in the number of overlapping sets. One simple fact is that if $A_p = 1$, then at least one element in $\{Z_{w'} : w' \neq w, p \in w'\}$ should be equal to 1. Based on this observation, one idea is to find disjoint sets of $Z_{w'}$'s which will result in $\prod_{p \in w} A_p = 1$ (this probability is easy to be computed using their independence), and then consider all the cases. We dervied an algorithm and coded a recursive function to implement it. In the function, we do something like a Depth First Search.

- Activate the whole containing the most parts (parts contained in this whole are thus turned on), and see which parts are left to be activated; if those remaining parts are contained by disjoint sets of wholes, compute the probability directly and go for the next whole containing the second most parts, and so on; if not, invoke the function recursively.
- Overlook wholes already activated and check if there are duplicate wholes which contain the same parts; if true (this happens a lot in GO), retain only one of them and record the number of duplicates;
- If there exists some part left to be activated which is not contained any left whole, stop the execution of the function.

One advantage of this algorithm is that we check if there are duplicate wholes and combine them at each recursion, which makes use of the local structure of the incidence matrix to the fullest extent and can save a lot on running time. In pseudo-code the algorithm is:

```
GetValue(p, wholes, parts)
  ## p is the Bernoulli parameter, wholes and parts are
  ## two lists storing the containing information
  if (there are duplicate wholes which contain the same parts)
      only retain one of them and record the number of duplicates;
  endif
  sort all the wholes according to the decreasing order of the number of parts
  they contain;
  for (i in 1:n )## length(wholes) = n
      remove wholes[0: i-1];
      turn on wholes[i] and check which parts are left to be activated;
      if (some part left to be activated is not included by any whole in wholes[i:n])
            break;
      endif
      if (sets of wholes containing these parts are disjoint)
            compute the probability, stored by prob[i];
      else (prob[i] = Getvalue(p, trimmed wholes, trimmed parts))
      ## trimmed wholes for ones containing parts left to be activated (parts in
      ## wholes[i] will be removed from the containing information) and
      ## trimmed parts for ones left to be activated
      endif
  endfor
  return sum(prob);
```

2 Estimating false-positive and true-positive rates

2.1 Type-2 diabetes example

Some R source to utilize the mixture-model fit from Morris et al 2012, Fig 1, in order to estimate the false positive (α) and true positive (γ) rates:

```
## Fig 1 in Morris et al 2012 shows a fitted mixture, applied to SNP-level Z scores,
## aiming to find out how many other T2D genes are out there

p1 <- 0.1465   ## proportion of involved genes
mu0 <- 0
mu1 <- 1.719
var0 <- 1
var1 <- 0.843

N <- 22000  ## guess on total number of genes in genome
x <- 77     ## number of transcripts significantly associated with disease, according to Morris et al
            ## at least, used in their pathway analysis...

## task 1: on the z-score scale, find an effective threshold that puts mass x/N beyond in the mixture model

mcdf <- function(z)
{
tmp1 <- (1-p1)*pnorm(z)
tmp2 <- p1*pnorm(z, mean=mu1, sd=sqrt(var1) )
return( tmp1+tmp2 )
}

foo <- function(zz)
{
dd <- (1-mcdf(zz)) - x/N
dd
}

cc <- ( uniroot( foo, interval=c(0,10) ) )$root
## cc    3.555

## task 2: on to role model parameters

## A = gene level activity indicator [truly affects T2D risk]
## Y = gene level call  [found in GWAS]

pY1 <- x/N
pA0 <- 1-p1

pY1gA0 <- pnorm( cc, lower.tail=FALSE )         ## alpha
pY1gA1 <- pnorm( cc, mean=mu1, sd=sqrt(var1), lower.tail=FALSE )  ## gamma

## 0.00019  alpha
## 0.02279  gamma

## check on implied FDR
pA0gY1 <- pY1gA0 * pA0/pY1  ## 0.046 ....reasonable
```

2.2 Influenza RNAi example

The 984 influenza-involved human genes reported in Hao *et al.* (2013) are those detected in the primary genome-wide screen of at least one of the four studies examined. From the detection model in that paper, we estimated α, the probability of detection in at least one study given the gene is not

truly involved in influenza, and $1-\gamma$, the probability of no detection in any study given the gene is truly involved. Using posterior mean estimates of the *metaflu* model parameters, and noting the key event is pattern '0000', we find:

$$\begin{aligned} \alpha &= P(\text{detect at least once}|\text{not involved}) = 1 - Prob(\text{"0000"}|I=0) = 0.01306943 \\ 1-\gamma &= P(\text{not detect at all}|\text{involved}) = Prob(\text{"0000"}|I=1) = 0.73197374. \end{aligned}$$

3 Algorithm for trimming activated sets

When reporting an inferred list of activated sets it is reasonable to remove from primary consideration those sets that are subsets of unions of larger sets in the list. This has a minimal effect in the T2D and RNAi case studies (1 set in each GO run), but it's helpful in general to see how many activated sets are effectively covering the gene list.

Pseudo-code

```
order sets inferred to be active,  from largest to smallest
loop over order
 if set is contained within union of previously retained sets
    then remove
 else retain
return: list of retained sets
```

R code

```
trimsets <- function(activeILP, xx)
{
## activeILP is vector of set id's
        ## xx is a nset x ngene incidence matrix
        size <- diag(t(xx[,activeILP])%*%xx[,activeILP])
        activeILPord <- names(sort(size, decreasing = T))
        activeILPsimple <- activeILPord[1]
        for (k in 2:length(activeILP))
                if(apply(as.matrix(xx[, activeILPsimple]), 1, max) %*%
                     xx[, activeILPord[k]] < size[activeILPord[k]])
                     activeILPsimple <- c(activeILPsimple, activeILPord[k])

        return(activeILPsimple)
}
```

Table S1: T2D: MFA-ILP results in Type-2 Diabetes example. Essentially the same as Table 3 (main paper), but with GO ID's included; coverage 26

GOID	Term (up to 40 characters)	Statistics	P.MFA	P.MGSA	FisherQ
GO:0001077	RNA polymerase II core promoter proximal	3/45	0.517	0.028	0.161
GO:0032024	positive regulation of insulin secretion	4/41	0.964	0.372	0.016
GO:0033138	positive regulation of peptidyl-serine p	2/35	0.537	0.096	0.756
GO:0046676	negative regulation of insulin secretion	4/23	0.996	0.201	0.003
GO:0006983	ER overload response	2/9	0.398	0.159	0.102
GO:0035774	positive regulation of insulin secretion	0/9	0.964	0.002	1
GO:0070365	hepatocyte differentiation	2/9	0.316	0.016	0.102
GO:0001714	endodermal cell fate specification	2/8	0.596	0.036	0.091
GO:0031017	exocrine pancreas development	3/8	0.946	0.6	0.003
GO:0032460	negative regulation of protein oligomeri	2/5	0.42	0.101	0.051
GO:0005638	lamin filament	2/5	0.79	0.4	0.051

Table S2: T2D: MGSA results in Type-2 Diabetes example. Similar to Table 3 (main paper), but the six gene sets shown are those inferred to be activated according to MGSA (P.MGSA $\geq$ .5); coverage 13

GOID	Term (up to 40 characters)	Statistics	P.MFA	P.MGSA	FisherQ
GO:0010506	regulation of autophagy	2/50	0	0.797	0.856
GO:0019915	lipid storage	3/49	0	0.708	0.192
GO:0017148	negative regulation of translation	2/46	0	0.552	0.856
GO:0070491	repressing transcription factor binding	3/30	0.053	0.823	0.074
GO:0031017	exocrine pancreas development	3/8	0.946	0.6	0.003

Table S3: T2D: Like Tables 3 and S1, but selected for small adjusted Fisher p-value; coverage 22

GOID	Term (up to 40 characters)	Statistics	P.MFA	P.MGSA	FisherQ
GO:0002793	positive regulation of peptide secretion	5/50	0	0.279	0.003
GO:0090277	positive regulation of peptide hormone s	5/49	0	0.337	0.003
GO:0046888	negative regulation of hormone secretion	4/43	0	0.149	0.018
GO:0033613	activating transcription factor binding	4/43	0.066	0.305	0.018
GO:0032024	positive regulation of insulin secretion	4/41	0.964	0.372	0.016
GO:0046323	glucose import	5/41	0	0.143	0.001
GO:0050994	regulation of lipid catabolic process	4/40	0	0.173	0.015
GO:0045913	positive regulation of carbohydrate meta	6/38	0	0.028	0
GO:0005978	glycogen biosynthetic process	5/37	0	0.016	0.001
GO:0009250	glucan biosynthetic process	5/37	0	0.02	0.001
GO:0046324	regulation of glucose import	4/37	0	0.017	0.012
GO:0010676	positive regulation of cellular carbohyd	6/35	0	0.027	0
GO:0032881	regulation of polysaccharide metabolic p	5/32	0	0.025	0
GO:0032885	regulation of polysaccharide biosyntheti	5/30	0	0.02	0
GO:0010907	positive regulation of glucose metabolic	6/28	0	0.02	0
GO:0070873	regulation of glycogen metabolic process	5/28	0	0.022	0
GO:0010828	positive regulation of glucose transport	4/27	0.002	0.019	0.004
GO:0045923	positive regulation of fatty acid metabo	4/27	0	0.013	0.004
GO:0002792	negative regulation of peptide secretion	4/26	0	0.171	0.003
GO:0005979	regulation of glycogen biosynthetic proc	5/26	0	0.022	0
GO:0010962	regulation of glucan biosynthetic proces	5/26	0	0.022	0
GO:0046326	positive regulation of glucose import	4/25	0.038	0.025	0.003
GO:0090278	negative regulation of peptide hormone s	4/25	0.004	0.208	0.003
GO:0046676	negative regulation of insulin secretion	4/23	0.996	0.201	0.003
GO:0006110	regulation of glycolysis	3/20	0	0.002	0.031
GO:0006111	regulation of gluconeogenesis	3/17	0	0.004	0.019
GO:0035987	endodermal cell differentiation	3/17	0	0.207	0.019
GO:0070875	positive regulation of glycogen metaboli	5/15	0	0.025	0
GO:0005159	insulin-like growth factor receptor bind	3/14	0.01	0.003	0.013
GO:0045725	positive regulation of glycogen biosynth	5/13	0.095	0.026	0
GO:0046321	positive regulation of fatty acid oxidat	3/11	0.001	0.014	0.006
GO:0031017	exocrine pancreas development	3/8	0.946	0.6	0.003

Table S4: T2D; GO terms with MFA marginal posterior activation probability exceeding 0.5. All of these sets are in the MAP estimate (Table S1).

GOID	Term (up to 40 characters)	Statistics	P.MFA	P.MGSA	FisherQ
GO:0001077	RNA polymerase II core promoter proximal	3/45	0.517	0.028	0.161
GO:0032024	positive regulation of insulin secretion	4/41	0.964	0.372	0.016
GO:0033138	positive regulation of peptidyl-serine p	2/35	0.537	0.096	0.756
GO:0046676	negative regulation of insulin secretion	4/23	0.996	0.201	0.003
GO:0035774	positive regulation of insulin secretion	0/9	0.964	0.002	1
GO:0001714	endodermal cell fate specification	2/8	0.596	0.036	0.091
GO:0031017	exocrine pancreas development	3/8	0.946	0.6	0.003
GO:0005638	lamin filament	2/5	0.79	0.4	0.051

Table S5: Influenza RNAi; GO terms found by MFA-ILP: coverage 245

GOID	Term (up to 40 characters)	Statistics	P.MFA	P.MGSA	FisherQ
GO:0015030	Cajal body	7/46	0.992	0.773	0.846
GO:0004715	non-membrane spanning protein tyrosine k	7/45	0.864	0.024	0.846
GO:0031526	brush border membrane	8/39	0.957	0.484	0.301
GO:0022627	cytosolic small ribosomal subunit	23/35	1	0.994	0
GO:0007094	mitotic cell cycle spindle assembly chec	5/33	0.216	0.097	0.882
GO:0032480	negative regulation of type I interferon	6/33	0.73	0.632	0.775
GO:0006730	one-carbon metabolic process	5/32	0.81	0.722	0.882
GO:0042147	retrograde transport, endosome to Golgi	5/32	0.819	0.76	0.882
GO:0032436	positive regulation of proteasomal ubiqu	10/31	1	0.103	0.007
GO:0004693	cyclin-dependent protein kinase activity	5/31	0.62	0.058	0.882
GO:0008542	visual learning	6/29	0.643	0.704	0.519
GO:0070888	E-box binding	5/27	0.597	0.061	0.846
GO:0043022	ribosome binding	7/26	0.731	0.03	0.152
GO:0000289	nuclear-transcribed mRNA poly(A) tail sh	5/26	0.841	0.915	0.841
GO:0048013	ephrin receptor signaling pathway	4/26	0.694	0.078	0.882
GO:0019003	GDP binding	4/26	0.632	0.029	0.882
GO:0006890	retrograde vesicle-mediated transport, G	10/24	0.971	0.909	0.002
GO:0005689	U12-type spliceosomal complex	8/24	1	0.834	0.026
GO:0005839	proteasome core complex	6/20	0.972	0.391	0.169
GO:0002089	lens morphogenesis in camera-type eye	5/19	0.696	0.48	0.404
GO:0010575	positive regulation vascular endothelial	4/18	0.266	0.39	0.841
GO:0004708	MAP kinase kinase activity	3/16	0.609	0.022	0.882
GO:0001502	cartilage condensation	5/16	0.997	0.971	0.25
GO:0051183	vitamin transporter activity	5/15	0.619	0.674	0.21
GO:0005852	eukaryotic translation initiation factor	9/14	1	0.992	0
GO:0042776	mitochondrial ATP synthesis coupled prot	4/13	0.933	0.007	0.417
GO:0051131	chaperone-mediated protein complex assem	4/13	0.436	0.083	0.417
GO:0008195	phosphatidate phosphatase activity	4/12	0.985	0.954	0.365
GO:0008536	Ran GTPase binding	3/12	0.813	0.65	0.882
GO:0019104	DNA N-glycosylase activity	3/11	0.618	0.054	0.846
GO:0015450	P-P-bond-hydrolysis-driven protein trans	5/11	0.612	0.688	0.065
GO:0005487	nucleocytoplasmic transporter activity	5/11	0.993	0.976	0.065
GO:0050321	tau-protein kinase activity	5/11	0.994	0.901	0.065
GO:0033179	proton-transporting V-type ATPase, V0 do	5/10	0.999	0.005	0.046
GO:0033180	proton-transporting V-type ATPase, V1 do	3/10	0.865	0.004	0.775
GO:0006600	creatine metabolic process	4/10	0.995	0.221	0.239
GO:0006465	signal peptide processing	3/10	0.856	0.569	0.775
GO:0008641	small protein activating enzyme activity	3/10	0.866	0.605	0.775
GO:0005838	proteasome regulatory particle	3/9	0.934	0.636	0.662
GO:0006013	mannose metabolic process	3/9	0.847	0.437	0.662
GO:0036002	pre-mRNA binding	3/9	0.832	0.369	0.662
GO:0006527	arginine catabolic process	3/8	0.514	0.074	0.519
GO:0019773	proteasome core complex, alpha-subunit c	1/8	0.976	0.004	0.882
GO:0009103	lipopolysaccharide biosynthetic process	3/8	0.909	0.135	0.519
GO:0019798	procollagen-proline dioxygenase activity	3/7	0.885	0.403	0.41
GO:0004703	G-protein coupled receptor kinase activi	3/7	0.943	0.892	0.41
GO:0002116	semaphorin receptor complex	4/7	0.545	0.144	0.079
GO:0030240	skeletal muscle thin filament assembly	2/5	0.486	0.265	0.846
GO:0007100	mitotic centrosome separation	2/5	0.509	0.013	0.846
GO:0005007	fibroblast growth factor-activated recep	3/5	0.866	0.42	0.22
GO:0017056	structural constituent of nuclear pore	3/5	0.461	0.381	0.22

Table S6: Influenza RNAi; GO terms found by MGSA; coverage 226

GOID	Term (up to 40 characters)	Statistics	P.MFA	P.MGSA	FisherQ
GO:0018107	peptidyl-threonine phosphorylation	12/49	0	0.886	0.019
GO:0032434	regulation of proteasomal ubiquitin-depe	12/48	0	0.897	0.017
GO:0000080	G1 phase of mitotic cell cycle	9/47	0.042	0.592	0.301
GO:0016469	proton-transporting two-sector ATPase co	14/47	0	0.963	0.002
GO:0015030	Cajal body	7/46	0.992	0.773	0.846
GO:0000245	spliceosomal complex assembly	6/37	0	0.539	0.846
GO:0007159	leukocyte cell-cell adhesion	5/37	0	0.504	0.882
GO:0050999	regulation of nitric-oxide synthase acti	8/35	0.009	0.836	0.197
GO:0008180	signalosome	6/35	0	0.947	0.841
GO:0022627	cytosolic small ribosomal subunit	23/35	1	0.994	0
GO:0004712	protein serine/threonine/tyrosine kinase	7/35	0.147	0.934	0.413
GO:0032480	negative regulation of type I interferon	6/33	0.73	0.632	0.775
GO:0006730	one-carbon metabolic process	5/32	0.81	0.722	0.882
GO:0042147	retrograde transport, endosome to Golgi	5/32	0.819	0.76	0.882
GO:0019843	rRNA binding	9/31	0.009	0.549	0.032
GO:0008542	visual learning	6/29	0.643	0.704	0.519
GO:0000289	nuclear-transcribed mRNA poly(A) tail sh	5/26	0.841	0.915	0.841
GO:2001236	regulation of extrinsic apoptotic signal	4/25	0	0.532	0.882
GO:0006890	retrograde vesicle-mediated transport, G	10/24	0.971	0.909	0.002
GO:0005689	U12-type spliceosomal complex	8/24	1	0.834	0.026
GO:0016775	phosphotransferase activity, nitrogenous	5/18	0	0.57	0.344
GO:0005003	ephrin receptor activity	4/17	0	0.835	0.775
GO:0001502	cartilage condensation	5/16	0.997	0.971	0.25
GO:0071526	semaphorin-plexin signaling pathway	6/16	0	0.599	0.065
GO:0051183	vitamin transporter activity	5/15	0.619	0.674	0.21
GO:0005852	eukaryotic translation initiation factor	9/14	1	0.992	0
GO:0090307	spindle assembly involved in mitosis	3/13	0.063	0.509	0.882
GO:0008195	phosphatidate phosphatase activity	4/12	0.985	0.954	0.365
GO:0008536	Ran GTPase binding	3/12	0.813	0.65	0.882
GO:0005487	nucleocytoplasmic transporter activity	5/11	0.993	0.976	0.065
GO:0015450	P-P-bond-hydrolysis-driven protein trans	5/11	0.612	0.688	0.065
GO:0050321	tau-protein kinase activity	5/11	0.994	0.901	0.065
GO:0006465	signal peptide processing	3/10	0.856	0.569	0.775
GO:0008641	small protein activating enzyme activity	3/10	0.866	0.605	0.775
GO:0005838	proteasome regulatory particle	3/9	0.934	0.636	0.662
GO:0004703	G-protein coupled receptor kinase activi	3/7	0.943	0.892	0.41

Table S7: Influenza RNAi; GO terms by Fisher's method; coverage 97

GOID	Term (up to 40 characters)	Statistics	P.MFA	P.MGSA	FisherQ
GO:0018107	peptidyl-threonine phosphorylation	12/49	0	0.886	0.019
GO:0032434	regulation of proteasomal ubiquitin-depe	12/48	0	0.897	0.017
GO:0016469	proton-transporting two-sector ATPase co	14/47	0	0.963	0.002
GO:0022627	cytosolic small ribosomal subunit	23/35	1	0.994	0
GO:0030532	small nuclear ribonucleoprotein complex	10/33	0	0.274	0.011
GO:0032436	positive regulation of proteasomal ubiqu	10/31	1	0.103	0.007
GO:0019843	rRNA binding	9/31	0.009	0.549	0.032
GO:0015988	energy coupled proton transport, against	10/30	0	0.015	0.006
GO:0015991	ATP hydrolysis coupled proton transport	10/30	0	0.005	0.006
GO:0006890	retrograde vesicle-mediated transport, G	10/24	0.971	0.909	0.002
GO:0005689	U12-type spliceosomal complex	8/24	1	0.834	0.026
GO:0033176	proton-transporting V-type ATPase comple	9/24	0	0.018	0.006
GO:0030137	COPI-coated vesicle	8/20	0	0.013	0.007
GO:0042274	ribosomal small subunit biogenesis	8/18	0	0.003	0.004
GO:0030663	COPI coated vesicle membrane	8/16	0	0.022	0.003
GO:0046933	hydrogen ion transporting ATP synthase a	7/15	0.003	0.007	0.007
GO:0005852	eukaryotic translation initiation factor	9/14	1	0.992	0
GO:0030126	COPI vesicle coat	8/14	0.03	0.043	0.001
GO:0035964	COPI-coated vesicle budding	7/13	0	0.006	0.003
GO:0048194	Golgi vesicle budding	7/13	0	0.003	0.003
GO:0048200	Golgi transport vesicle coating	7/13	0	0.006	0.003
GO:0048205	COPI coating of Golgi vesicle	7/13	0	0.006	0.003
GO:0033179	proton-transporting V-type ATPase, V0 do	5/10	0.999	0.005	0.046
GO:0000028	ribosomal small subunit assembly	4/6	0.014	0.004	0.046

Table S8: Influenza RNAi; GO terms with MFA marginal posterior activation probability exceeding 0.5 . Three sets (GO:0046330, GO:0005719, GO:0051233) are not in the MAP estimate (Table S5).

GOID	Term (up to 40 characters)	Statistics	P.MFA	P.MGSA	FisherQ
GO:0046330	positive regulation of JNK cascade	9/46	0.714	0.191	0.27
GO:0015030	Cajal body	7/46	0.992	0.773	0.846
GO:0004715	non-membrane spanning protein tyrosine k	7/45	0.864	0.024	0.846
GO:0031526	brush border membrane	8/39	0.957	0.484	0.301
GO:0022627	cytosolic small ribosomal subunit	23/35	1	0.994	0
GO:0032480	negative regulation of type I interferon	6/33	0.73	0.632	0.775
GO:0006730	one-carbon metabolic process	5/32	0.81	0.722	0.882
GO:0042147	retrograde transport, endosome to Golgi	5/32	0.819	0.76	0.882
GO:0032436	positive regulation of proteasomal ubiqu	10/31	1	0.103	0.007
GO:0004693	cyclin-dependent protein kinase activity	5/31	0.62	0.058	0.882
GO:0008542	visual learning	6/29	0.643	0.704	0.519
GO:0070888	E-box binding	5/27	0.597	0.061	0.846
GO:0000289	nuclear-transcribed mRNA poly(A) tail sh	5/26	0.841	0.915	0.841
GO:0048013	ephrin receptor signaling pathway	4/26	0.694	0.078	0.882
GO:0019003	GDP binding	4/26	0.632	0.029	0.882
GO:0043022	ribosome binding	7/26	0.731	0.03	0.152
GO:0006890	retrograde vesicle-mediated transport, G	10/24	0.971	0.909	0.002
GO:0005689	U12-type spliceosomal complex	8/24	1	0.834	0.026
GO:0005839	proteasome core complex	6/20	0.972	0.391	0.169
GO:0002089	lens morphogenesis in camera-type eye	5/19	0.696	0.48	0.404
GO:0001502	cartilage condensation	5/16	0.997	0.971	0.25
GO:0004708	MAP kinase kinase activity	3/16	0.609	0.022	0.882
GO:0051183	vitamin transporter activity	5/15	0.619	0.674	0.21
GO:0005852	eukaryotic translation initiation factor	9/14	1	0.992	0
GO:0042776	mitochondrial ATP synthesis coupled prot	4/13	0.933	0.007	0.417
GO:0051233	spindle midzone	4/12	0.708	0.16	0.365
GO:0008195	phosphatidate phosphatase activity	4/12	0.985	0.954	0.365
GO:0008536	Ran GTPase binding	3/12	0.813	0.65	0.882
GO:0005487	nucleocytoplasmic transporter activity	5/11	0.993	0.976	0.065
GO:0015450	P-P-bond-hydrolysis-driven protein trans	5/11	0.612	0.688	0.065
GO:0019104	DNA N-glycosylase activity	3/11	0.618	0.054	0.846
GO:0050321	tau-protein kinase activity	5/11	0.994	0.901	0.065
GO:0006465	signal peptide processing	3/10	0.856	0.569	0.775
GO:0006600	creatine metabolic process	4/10	0.995	0.221	0.239
GO:0005719	nuclear euchromatin	3/10	0.591	0.367	0.775
GO:0033179	proton-transporting V-type ATPase, V0 do	5/10	0.999	0.005	0.046
GO:0033180	proton-transporting V-type ATPase, V1 do	3/10	0.865	0.004	0.775
GO:0008641	small protein activating enzyme activity	3/10	0.866	0.605	0.775
GO:0006013	mannose metabolic process	3/9	0.847	0.437	0.662
GO:0005838	proteasome regulatory particle	3/9	0.934	0.636	0.662
GO:0036002	pre-mRNA binding	3/9	0.832	0.369	0.662
GO:0006527	arginine catabolic process	3/8	0.514	0.074	0.519
GO:0009103	lipopolysaccharide biosynthetic process	3/8	0.909	0.135	0.519
GO:0019773	proteasome core complex, alpha-subunit c	1/8	0.976	0.004	0.882
GO:0002116	semaphorin receptor complex	4/7	0.545	0.144	0.079
GO:0004703	G-protein coupled receptor kinase activi	3/7	0.943	0.892	0.41
GO:0019798	procollagen-proline dioxygenase activity	3/7	0.885	0.403	0.41
GO:0007100	mitotic centrosome separation	2/5	0.509	0.013	0.846
GO:0005007	fibroblast growth factor-activated recep	3/5	0.866	0.42	0.22

Table S9: Coverage, miscoverage, and other statistics for three methods in two case studies. Recall y_g indicates whether or not gene g is on the observed gene list; $\hat{A}_g$ is whether or not the gene is in a set that is inferred to be active. Entries (except last column) are numbers of genes satisfying the condition. In each case and gene-set system, the first two numerical columns add to a constant, as do the next two columns. The final column indicates the number of sets inferred to be active, after trimming. MFA-ILP delivers relatively high coverage and low mis-coverage in all cases.

		coverage		mis-coverage		
		y_g=1, $\hat{A}_g = 1$	y_g=1, $\hat{A}_g = 0$	y_g=0, $\hat{A}_g = 1$	y_g=0, $\hat{A}_g = 0$	# of sets
T2D, GO	Fisher	22	36	265	10303	16
	MGSA	13	45	169	10399	5
	MFA-ILP	26	32	156	10412	10
RNAi, GO	Fisher	97	586	206	8749	13
	MGSA	226	457	634	8321	36
	MFA-ILP	245	438	635	8320	50
T2D, KEGG	Fisher	12	8	55	1846	2
	MGSA	20	0	174	1727	6
	MFA-ILP	20	0	174	1727	6
RNAi, KEGG	Fisher	32	132	71	1686	3
	MGSA	111	53	522	1235	22
	MFA-ILP	108	56	497	1260	21

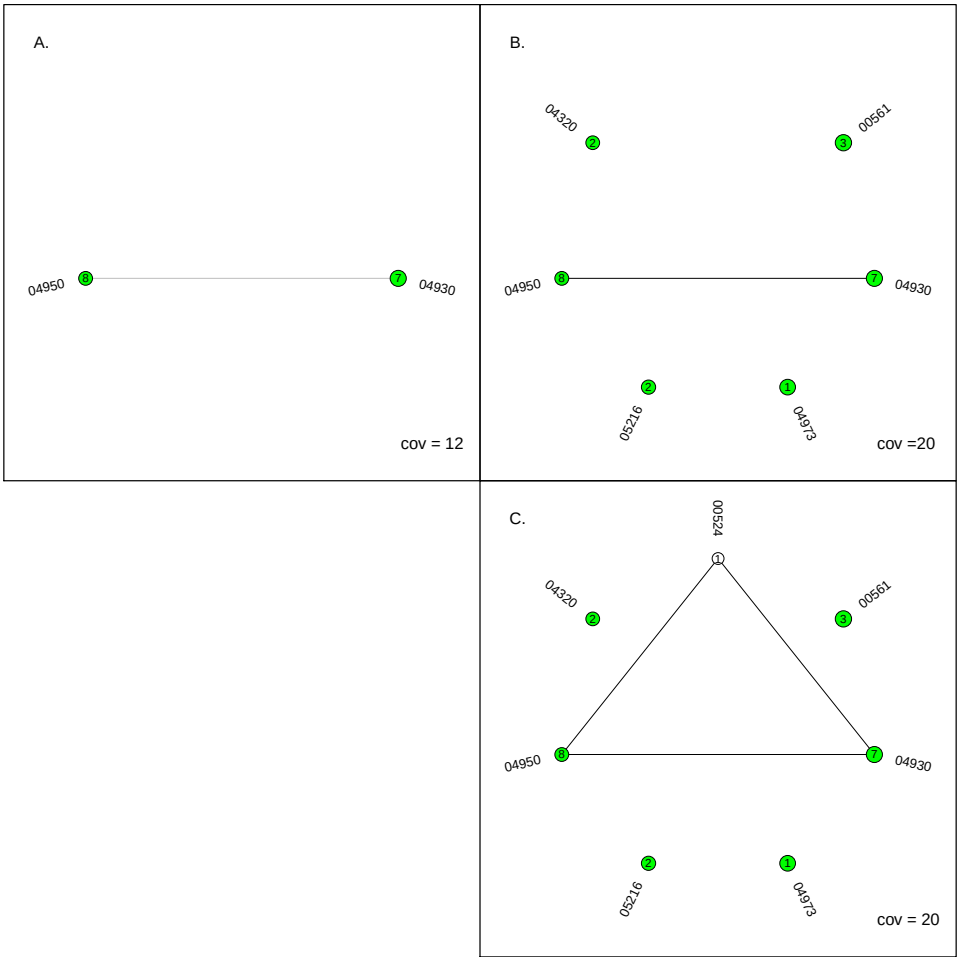

Figure S1: KEGG pathways identified by three methods (A, Fisher's test; B, MGSA; C, MFA-ILP) as activated to explain the type 2 diabetes associated genes. Layout as in Figures 2 and 3 (main paper). There is very close agreement between MGSA and MFA-ILP.

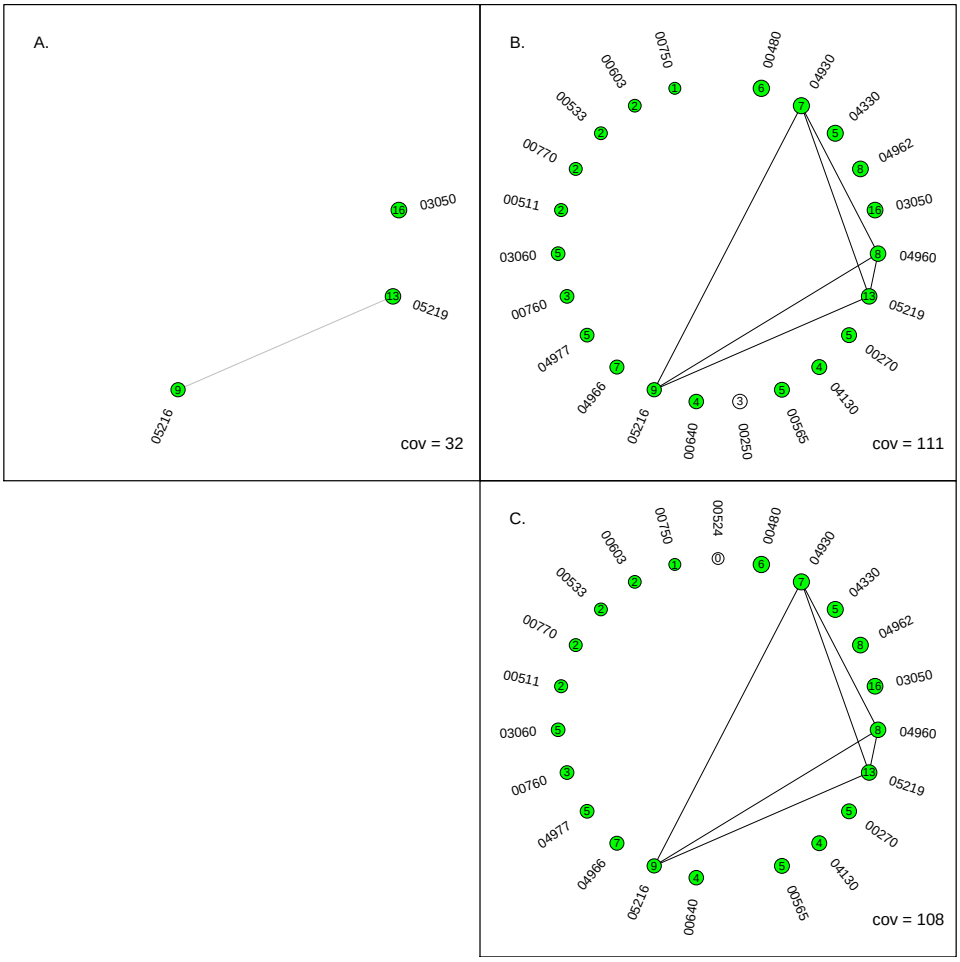

Figure S2: KEGG pathways identified by three methods (A, Fisher's test; B, MGSA; C, MFA-ILP) as activated to explain the influenza RNAi genes. Layout as in Figures 2 and 3 (main paper). There is very close agreement between MGSA and MFA-ILP.