Supplements

**Supplement Figure S1.** Reproducible prediction of PIN-reprioritization using higher confidence protein interactions (Combined Scores>900). Page 2

**Supplement Figure S2.** Consistent prediction of PIN-reprioritization of GWAS-ranked genes including host genes and the nearest genes of intergenic SNPs. Page 3

**Supplement Figure S3.** A control study of KEGG pathways reprioritization of GWAS SNPs performs similarly or slightly better than GWAS p-value prioritization in discovering known Trait-Associated SNPs from the independent Gold Standard, however it does not outperform SPAN. Page 4

**Supplement Table S1.** 21 genes of the optimal SPAN model. Page 5

**Supplement Table S2.** Gold standard for T2D of FUSION. Page 6

**Supplement Table S3.** Gold standard for T2D of WTCCC. Page 7

**Supplement Table S4.** Gold standard for Crohn's disease of IBDGC. Page 8

**Supplement Table S5.** 97 genes associated to Type 2 Diabetes reported in the Online Mendelian Inheritance in Man. Page 9-10

**Supplement Table S6.** Validated T2D genes are enriched in the Optimal SPAN Model of T2D: Possible role of prioritized host genes in glucose homeostasis and diabetes mellitus. Page 11-14

**Supplement Table S7.** The empirical frequency of SPAN frequency (P-value) of the 21 genes in Figure 4. Page 15

**Supplement Table S8.** Edgetic P-value of interactions and evidence. Page 16

**Supplement Table S9.** Edgetic P-values and supporting evidence for the direct interaction of the 21 PIN-prioritized genes with 17 known T2D genes. Page 17

**Supplement Table S10.** Table of Abbreviations, Terms, and key concepts. Page 18

**Supplement Methods.** Page 19-23
Supplement Figure S1. Reproducible Prediction of PIN-reprioritization Using Higher Confidence Protein Interactions (Combined Scores $>900$). We conducted PIN-reprioritization using protein-protein interactions with higher confidence evidence which was compiled by collecting all protein-protein interactions having a combined score $>900$ in STRING version 6.3 and 8.2. PIN-reprioritization using this higher confidence STRING dataset reproduced the predictions made using the original STRING dataset used in our analysis (STRING version 6.3 and 8.2 without text mining, not restricted for confidence score).
Supplement Figure S2. Consistent prediction of PIN-reprioritization of GWAS-ranked genes including host genes and the nearest genes of intergenic SNPs. We conducted a PIN-reprioritization of GWAS-ranked genes that including both host genes of intragenic SNPs and the nearest genes of intergenic SNPs. The PIN-reprioritization was performed using the same STRING database used for our original analysis. To match the nearest gene to intergenic SNPs, we downloaded three tables, hgncXref.txt.gz, knownGene.txt.gz, and snp129.txt.gz from the UCSC Genome browser (http://hgdownload.cse.ucsc.edu/goldenPath/hg18/database) on Sept. 11 2011. Using the genomic coordinates of transcription start and end sites of a gene, we calculated the physical distance between SNPs and the genes located directly upstream and downstream of the SNP. Between the up and downstream gene, the gene with the shorter distance was assigned to the SNP as the “nearest gene”. The PIN-reprioritization of this expanded set of GWAS-ranked genes (including both host genes and the nearest genes of intergenic SNPs) showed similar or slightly reduced ranges of odds ratios as compared to the analysis only considering host genes of intragenic SNPs. Interestingly, the maximal odds ratios were consistently observed between GWAS-ranked sets of 500 and 600 for both this analysis and the original analysis as shown in Figure 4 for only GWAS-ranked host genes.
Supplement Figure S3. A control study of KEGG pathways reprioritization of GWAS SNPs performs similarly or slightly better than GWAS p-value prioritization in discovering known Trait-Associated SNPs from the independent Gold Standard, however it does not outperform SPAN. We and others have previously reported that pathway enrichment or genesets can uncover SNPs buried in GWAS not detected in the initial study [1, 2]. In order to compare the accuracy of the SPAN algorithm proposed in this manuscript to that of pathway enrichment in discovering SNPs buried in GWAS, we utilized all pathways from KEGG and systematically verified the pathway enrichment at each host gene cutoff. SNPs of host genes uncovered by significant pathways at each FDR threshold were selected and an odds ratio was computed using the "Reference Gold Standard" GS (Methods) and Fisher Exact Test. The X-axis shows host gene cutoffs and the y-axis shows the odds ratios of recapitulating known trait-associated SNPs with respect to various host gene cutoffs and enrichment significance denoted by false discovery rate cutoffs (FDR line colors). In summary, the KEGG enrichment prioritization was slightly better than the GWAS in one of the two studies only, thus not reproducible across datasets. In contrast, the SPAN protein interaction network reprioritization method, shown in Figure 4, robustly reproduced much higher odds ratio than KEGG or GWAS methods.

References:
1 Lee Y, Li J, Gamazon E, Chen JL, Tikhomirov A, Cox NJ, Lussier YA. Biomolecular systems of disease buried across multiple GWAS unveiled by information theory and ontology. AMIA Summits Transl Sci Proc. 2010 Mar 1;2010:31-5.
2 Province MA, Borecki IB. Gathering the gold dust: methods for assessing the aggregate impact of small effect genes in genomic scans. Pac Symp Biocomput 2008:190-200.
Supplement Table S1. 21 genes of the optimal SPAN model. We selected the model at 600 GWAS-ranked host genes with a frequency ≤0.1% for our optimal network since it has the highest odds ratio with a P-value=0.00059 for FUSION and a P-value=0.00073 for WTCCC. This model contains 12 host genes (corresponding to 12 SNPs) from FUSION and 10 host genes (corresponding to 10 SNPs) from WTCCC and only one host gene, KCNJ11 (rs5215), is common between both studies. Thus the combined network is comprised of 21 distinct genes (Figure 5 and 6). Five genes are 1st interactors of gold standard according to our protein interaction dataset (Methods). Furthermore, 7 and 6 genes have topological properties of bottleneck and hubness, respectively. 1 Intragenic SNPs were mapped to host genes according to dbSNP annotation and 2 17 genes of union of gold standard for FUSION and gold standard for WTCCC (See Methods). 3 FUSION and 4 WTCCC

| rs          | GWAS | Gene          | SNP annotation 1 | Protein Interaction Properties |
|-------------|------|---------------|------------------|-------------------------------|
| Function    | Chromosome | 1st interactor of GS 2 | Bottleness | Hubness |
| rs17184300  | F^3   | ARG1          | near-gene-3      | chr6                          |                       |                       |
| rs11217854  | W^4   | ARHGEF12      | intron           | chr11                         |                       |                       |
| rs6578410   | F     | ART1          | near-gene-5      | chr11                         |                       |                       |
| rs7359414   | F     | AXIN1         | intron           | chr16                         |                       |                       |
| rs2056975   | W     | CDC42         | intron           | chr1                          |                       |                       |
| rs4958228   | F     | CDKL3         | intron           | chr5                          |                       |                       |
| rs2505639   | W     | CREM          | intron           | chr10                         |                       |                       |
| rs1033583   | F     | DLL1          | utr-3            | chr6                          |                       |                       |
| rs664893    | W     | IL28A         | near-gene-5      | chr19                         |                       |                       |
| rs1130183   | F     | KCNJ10        | missense         | chr1                          |                       |                       |
| rs5215      | W,F   | KCNJ11        | reference        | chr11                         |                       |                       |
| rs2895      | W     | LFNG          | utr-3            | chr7                          |                       |                       |
| rs726501    | F     | MAP3K1        | intron           | chr5                          |                       |                       |
| rs6525591   | F     | PIN4          | intron           | chrX                          |                       |                       |
| rs3796224   | W     | PROK2         | intron           | chr3                          |                       |                       |
| rs165598    | F     | SNAP29        | intron           | chr22                         |                       |                       |
| rs17849165  | W     | TACR3         | intron           | chr4                          |                       |                       |
| rs17136481  | W     | TRAP1         | intron           | chr16                         |                       |                       |
| rs254456    | F     | TRIM7         | near-gene-3      | chr5                          |                       |                       |
| rs10927875  | F     | ZBTB17        | intron           | chr1                          |                       |                       |
| rs4803674   | W     | ZNF284        | intron           | chr19                         |                       |                       |

1 Intragenic SNPs were mapped to host genes according to dbSNP annotation and 
2 17 genes of union of gold standard for FUSION and gold standard for WTCCC (See Methods). 3 FUSION and 4 WTCCC
**Supplement Table S2. Gold standard for T2D of FUSION**  
To construct a gold standard for T2D SNPs from this data, we extracted 76 SNPs which are reported with either “Type 2 diabetes and other traits”, “Type 2 diabetes and 6 quantitative traits” or “Type 2 diabetes” in the Disease/Trait column. Among 76 SNPs, 42 are intragenic SNPs and correspond to 22 host genes according to dbSNP annotations. Finally, we selected a list of SNPs containing 11 with 10 corresponding host genes which are contained in FUSION (Illumina Infinium™ II Human Hap300 BeadChips v.1.0) platforms as well as in the protein interaction dataset. These sets of SNPs were used to assess the accuracy of network models curated for FUSION. 1 Intragenic SNPs were mapped to host genes according to dbSNP annotation (Methods). 2,3 NHGRI SNP's rank in all GWAS-ranked SNPs. 3 NHGRI SNP's rank in GWAS-ranked intragenic SNPs. 4 Host gene's rank in GWAS-ranked host genes.

| rs         | Host gene | SNP Rank in all SNPs of platform | SNP Rank in intragenic SNP of platform | Host gene rank in platform | PubMed ID          |
|------------|-----------|----------------------------------|---------------------------------------|---------------------------|--------------------|
| rs1470579  | IGF2BP2   | 70                               | 31                                    | 28                        | 20581827           |
| rs7901695  | TCF7L2    | 367                              | 155                                   | 2                         | 17463249           |
| rs7756992  | CDKAL1    | 739                              | 300                                   | 83                        | 17460697           |
| rs5215     | KCNJ11    | 1323                             | 552                                   | 400                       | 18372903, 17463249 |
| rs7578597  | THADA     | 2392                             | 987                                   | 207                       | 18372903           |
| rs4712523  | CDKAL1    | 4514                             | 1903                                  | 83                        | 19734900, 19401414 |
| rs8042680  | PRC1      | 14320                            | 6105                                  | 1922                      | 20581827           |
| rs896854   | TP53INP1  | 23085                            | 9970                                  | 4050                      | 20581827           |
| rs4689388  | WFS1      | 106957                           | 46413                                 | 5391                      | 19734900           |
| rs391300   | SRR       | 217389                           | 94420                                 | 14916                     | 20174558           |
| rs2237892  | KCNQ1     | 271260                           | 118014                                | 421                       | 19401414, 18711367 |
Supplement Table S3. Gold standard for T2D of WTCCC. To construct a gold standard for T2D SNPs from this data, we extracted 76 SNPs which are reported with either “Type 2 diabetes and other traits”, “Type 2 diabetes and 6 quantitative traits” or “Type 2 diabetes” in the Disease/Trait column. Among 76 SNPs, 42 are intragenic SNPs and correspond to 22 host genes according to dbSNP annotations. Finally, we selected a list of SNPs containing 10 with 8 corresponding host genes which are contained in WTCCC platforms as well as in the protein interaction dataset. These sets of SNPs were used to assess the accuracy of network models curated for WTCCC. Intragenic SNPs were mapped to host genes according to dbSNP annotation (Methods). Affymetrix GeneChip 500K. NHGRI SNP’s rank in all GWAS-ranked SNPs. NHGRI SNP’s rank in GWAS-ranked intragenic SNPs. Host gene’s rank in GWAS-ranked host genes.

| rs       | Host gene | SNP Rank in all SNPs of platform | SNP Rank in intragenic SNP of platform | Host gene rank in platform | PubMed ID               |
|----------|-----------|----------------------------------|----------------------------------------|---------------------------|-------------------------|
| rs7901695| TCF7L2     | 11                               | 6                                      | 5                         | 17463249                |
| rs7593730| RBMS1      | 52                               | 38                                     | 17                        | 20418489                |
| rs10946398| CDKAL1     | 97                               | 63                                     | 14                        | 19056611, 17463249      |
| rs7754840| CDKAL1     | 118                              | 71                                     | 14                        | 17463246, 17463248      |
| rs864745 | JAZF1      | 236                              | 125                                    | 74                        | 18372903                |
| rs1801282| PPARG      | 955                              | 446                                    | 248                       | 17463246, 17463248, 17463249 |
| rs5215   | KCNJ11     | 967                              | 452                                    | 260                       | 18372903, 17463249      |
| rs4402960| IGF2BP2    | 1210                             | 558                                    | 324                       | 19401414, 18372903, 17463246, 17463248, 17463249, 20581827 |
| rs1470579| IGF2BP2    | 1759                             | 799                                    | 324                       | 20581827                |
| rs10923931| NOTCH2     | 3376                             | 1469                                   | 774                       | 18372903                |
Supplement Table S4. Gold standard for Crohn's disease of IBDGC  To construct a gold standard for Crohn's and inflammatory bowel disease from the NHGRI catalog (http://www.genome.gov/gwastudies/), we extracted 81 SNPs which are reported with either “Crohn's disease”, “Inflammatory bowel disease”, “Crohn's disease and sarcoidosis (combined)”, or “Inflammatory bowel disease (early onset)” in the Disease/Trait column. Among the 81 SNPs identified, 40 are intragenic and correspond to 23 host genes according to dbSNP annotations. Finally, we selected all 20 SNPs (15 corresponding host genes) that are present in the IBDGC GWAS platform (Illumina Infinium™ II Human Hap300 BeadChips v.1.0) as well as in the protein interaction dataset. These 20 SNPs were used to assess the accuracy of our Crohn’s network models. Intragenic SNPs were mapped to host genes according to dbSNP annotation (Methods). HNHGRI SNP's rank in all GWAS-ranked SNPs. NHGRI SNP's rank in GWAS-ranked intragenic SNPs. Host gene's rank in GWAS-ranked host genes.

| rs     | Host gene | SNP Rank in all SNPs of platform | SNP Rank in intragenic SNP of platform | Host gene rank in platform | PubMed ID              |
|--------|-----------|----------------------------------|----------------------------------------|---------------------------|------------------------|
| rs5743289 | NOD2      | 14                               | 12                                     | 1                         | 18758464, 17804789, 17447842 |
| rs1343151 | IL23R     | 4                                | 4                                      | 2                         | 17804789               |
| rs10889677 | IL23R     | 7                                | 7                                      | 2                         | 17804789               |
| rs2201841 | IL23R     | 8                                | 8                                      | 2                         | 17804789               |
| rs111465804 | IL23R    | 9                                | 9                                      | 2                         | 20570966, 18587394, 17804789 |
| rs1004819 | IL23R     | 11                               | 10                                     | 2                         | 17804789               |
| rs2064689 | IL23R     | 55                               | 30                                     | 2                         | 17804789               |
| rs1250550 | ZMIZ1     | 60                               | 34                                     | 21                        | 19915574               |
| rs11190140 | NKX2-3    | 12693                            | 5564                                   | 64                        | 18587394               |
| rs2274910 | ITLN1     | 274                              | 114                                    | 89                        | 18587394               |
| rs504963  | FUT2      | 949                              | 406                                    | 317                       | 20570966               |
| rs2301436 | FGFR1OP   | 1960                             | 877                                    | 568                       | 20570966, 18587394     |
| rs2476601 | PTPN22    | 3174                             | 1416                                   | 694                       | 18587394               |
| rs6908425 | CDKAL1    | 20791                            | 9169                                   | 1059                      | 18587394               |
| rs3764147 | C13orf31  | 8554                             | 3722                                   | 2093                      | 18587394               |
| rs6478109 | TNFSF15   | 21748                            | 9573                                   | 4170                      | 18758464               |
| rs2315008 | ZGAPAT    | 28504                            | 12629                                  | 4774                      | 18758464               |
| rs3197999 | MST1      | 27368                            | 12108                                  | 4884                      | 18587394               |
| rs8049439 | ATXN2L    | 66889                            | 29441                                  | 8570                      | 19915574               |
| rs744166  | STAT3     | 87512                            | 38356                                  | 9921                      | 18587394               |
Supplement Table S5. 97 genes associated to Type 2 Diabetes reported in the Online Mendelian Inheritance in Man (OMIM[MIM#12583 - "NIDDM" ; 12/2010])

| PubMed ID | Gene symbol | Availability in protein interaction dataset | PubMed ID | Gene symbol | Availability in protein interaction dataset |
|-----------|-------------|--------------------------------------------|-----------|-------------|--------------------------------------------|
| 6833      | ABCC8       | Yes                                        | 11183     | MAP4K5      | Yes                                        |
| 208       | AKT2        | Yes                                        | 9479      | MAPK8IP1    | Yes                                        |
| 11132     | CAPN10      | Yes                                        | 10573     | MRPL28      | Yes                                        |
| 6347      | CCL2        | Yes                                        | 4681      | NBL1        | Yes                                        |
| 1231      | CCR2        | Yes                                        | 4536      | ND2         | Yes                                        |
| 1026      | CDKN1A      | Yes                                        | 4540      | ND5         | Yes                                        |
| 1029      | CDKN2A      | Yes                                        | 4760      | NEUROD1     | Yes                                        |
| 1030      | CDKN2B      | Yes                                        | 4790      | NFKB1       | Yes                                        |
| 10664     | CTCF        | Yes                                        | 4813      | NIDDM2      | Yes                                        |
| 6387      | CXCL12      | Yes                                        | 50982     | NIDDM3      | Yes                                        |
| 27065     | D4S234E     | Yes                                        | 4842      | NOS1        | Yes                                        |
| 1756      | DMD         | Yes                                        | 4843      | NOS2A       | Yes                                        |
| 8894      | EIF2S2      | Yes                                        | 29107     | NXT1        | Yes                                        |
| 1968      | EIF2S3      | Yes                                        | 5078      | PAX4        | Yes                                        |
| 79071     | ELOVL6      | Yes                                        | 5465      | PPARA       | Yes                                        |
| 2053      | EPHX2       | Yes                                        | 5468      | PPARG       | Yes                                        |
| 51013     | EXOSC1      | Yes                                        | 5581      | PRKCE       | Yes                                        |
| 2246      | FGF1        | Yes                                        | 5770      | PTPN1       | Yes                                        |
| 2255      | FGF10       | Yes                                        | 56729     | RETN        | Yes                                        |
| 2258      | FGF13       | Yes                                        | 6462      | SHBG        | Yes                                        |
| 2247      | FGF2        | Yes                                        | 6514      | SLC2A2      | Yes                                        |
| 2249      | FGF4        | Yes                                        | 6517      | SLC2A4      | Yes                                        |
| 2250      | FGF5        | Yes                                        | 169026    | SLC30A8     | Yes                                        |
| 2252      | FGF7        | Yes                                        | 10923     | SUB1        | Yes                                        |
| 2260      | FGFRI       | Yes                                        | 6927      | TCF1        | Yes                                        |
| 79068     | FTO         | Yes                                        | 6928      | TCF2        | Yes                                        |
| 2572      | GAD2        | Yes                                        | 6934      | TCF7L2      | Yes                                        |
| 2645      | GCK         | Yes                                        | 7021      | TFAP2B      | Yes                                        |
| 2646      | GCKR        | Yes                                        | 8797      | TNFRSF10A   | Yes                                        |
| 3077      | HFE         | Yes                                        | 8718      | TNFRSF25    | Yes                                        |
| 3087      | HHEX        | Yes                                        | 200186    | TORC2       | Yes                                        |
| 3119      | HLA-DQB1    | Yes                                        | 7103      | TSPAN8      | Yes                                        |
| 3159      | HMGA1       | Yes                                        | 7439      | VMD2        | Yes                                        |
| 3172      | HNF4A       | Yes                                        | 7466      | WFS1        | Yes                                        |
| 3416      | IDE         | Yes                                        | 9370      | ADIPOQ      | No                                         |
| 3551      | IKBP       | Yes                                        | 51129     | ANGPTL4     | No                                         |
| 3569      | IL6         | Yes                                        | 54901     | CDKAL1      | No                                         |
| 3630      | INS         | Yes                                        | 1965      | EIF2S1      | No                                         |
| 3643      | INSR        | Yes                                        | 5167      | ENPP1       | No                                         |
| 3651      | IPF1        | Yes                                        | 2820      | GPD2        | No                                         |
|    | Gene  | Expression  |
|----|-------|-------------|
| 3667 | IRS1  | Yes         |
| 8660 | IRS2  | Yes         |
| 3767 | KCNJ11| Yes         |
| 3784 | KCNQ1 | Yes         |
| 3832 | KIF11 | Yes         |
| 3898 | LAD1  | Yes         |
| 10660| LBX1  | Yes         |
| 3952 | LEP   | Yes         |
| 5871 | MAP4K2| Yes         |
| 3727 | KCNJ15| No          |
| 8473 | OGT   | No          |
| 8050 | PDHX  | No          |
| 57804| POLD4 | No          |
| 56655| POLE4 | No          |
| 5506 | PPP1R3A| No         |
| 9317 | PTER  | No          |
| 9338 | TCEAL1| No          |
Supplement Table S6. Validated T2D genes are enriched in the Optimal SPAN Model of T2D: Possible role of prioritized host genes in glucose homeostasis and diabetes mellitus. Since the gold standard was derived from the NHGRI catalog, its capacity for evaluation is limited by the breadth of available GWAS results. To expand the evaluation and assess the robustness of our optimal T2D network model, we utilized two independent resources, Online Mendelian Inheritance in Man (OMIM) and Ingenuity Pathway Analysis (IPA, www.ingenuity.com), and conducted a review of literature of canonical pathways (Figure 5E) to provide supplement validation unconstrained by GWAS. We reviewed the literature to provide evidence in support of the association between T2D and the optimal SPAN-derived network illustrated in Figure 5 that comprises the host genes of re-prioritized SNPs from two T2D GWAS. Each host gene was first entered into the Gene Cards browser (http://www.genecards.org/) where the disorders section, listing Novoseek Disease relationships, was curated for T2D and related disorders. The annotated Pub med IDs (PMID) were examined for true linkage between the disorder and the gene of interest. If no conclusive references were presented we extended the search to PubMed (http://www.ncbi.nlm.nih.gov/pubmed/). As a final verification, the genes’ canonical pathways and biological mechanisms relevant to T2D in KEGG, GO, Ingenuity Pathway Analysis (IPA), and Reactome [1] were searched. * Figure 5C, T2D related gene in yellow, or Glucose Homeostasis related gene in mauve).

| Host Gene | Rationale * | Type of Evidence | References |
|-----------|-------------|------------------|------------|
| ARHGEF12  | Variant of the LARG gene (ARGEF12 alias) was found to be associated with increased insulin action [2] | Genetic sequencing study | PubMed [2] |
| ART1      | ART2.2 (ART1 alias) inhibition in NOD.cd38 mice allowed for restoration of natural killer cell population that, when activated, were able to inhibit Type I Diabetes development [3]. This evidence is listed because T2D GWAS may contain SNPs of T1D due to the ambiguous clinical diagnosis of some diabetic individuals. | In vivo (mice) | PubMed |
| AXIN1     | AXIN-1 is an Inhibitor of the WNT signaling pathway which has been shown to be linked to T2D development[4] | Genetic population study | PubMed [4,5] |
|           | WNT signaling also shown to reduce pancreatic β-cell growth and impair glucose tolerance in mice [5] | In vivo (mice) | PubMed [5] |
| CREM      | CREM splicing variant effectively represses insulin gene transcription[6] | In vivo (rats) | PubMed |
| KCNJ10    | Found to be associated to T2D risk locus via linkage disequilibrium, however it may not be | Genetic study | PubMed |
| Gene          | Description                                                                 | Method            | Database     |
|--------------|-----------------------------------------------------------------------------|-------------------|--------------|
| KCNJ11       | a causative heritable factor of T2D [7]                                      |                   | T2D          |
| T2D Ingenuity pathway | NHGRI GWAS Compendium annotates one of its SNP, serves as a gold standard gene for this study | Curation          | IPA          |
| Neonatal T2D[8]      | Curation                                                                      | Sequencing study  | PubMed       |
| T2D development[9]   | Case control data meta analysis                                             |                   | PubMed       |
| T2D pathway        | Congenital Hyperinsulinemia [10,11]                                         |                   | PubMed       |
| MAP3K1          | Inhibits cAMP-induced insulin transcription in pancreatic β-cells [12]       | In vivo (mice)    | PubMed [12]  |
| MIM: 600982     | Integrates cellular response to insulin                                       |                   | IPA          |
| PROK2           | Neurologically inhibits food intake [13], related to obesity (a T2D related disorder) | In vivo (rats)    | PubMed       |
| SNAP29          | Insulin secretory defect associated to protein family                        | In vivo (rats)    |             |
| TACR3           | Statistical association to T2D via meta analysis of WTCCC GWAS [14]          | Statistical association to T2D | PubMed |
| TRAP1           | TRAP-1 is a Ligand of the TNF-α receptor, TNF-α is part of the Type II Diabetes Mellitus KEGG pathway [15] | KEGG             |             |
| TRIM7           | GNIP (TRIM7 alias) interacts with Glycogenin by increasing the rate of reaction, glycogen metabolism is significantly altered in diabetes and glycogenin may be involved in the genetic portion of T2D [17] | Protein interaction study | PubMed    |
| T2D sub-pathway [18], activation of chaperone genes by XBP1 (s) | **curation** | | PubMed |
| ARG1           | NONE                                                                         |                   |             |
| CDC42          | NONE                                                                         |                   |             |
| CDKL3          | NONE                                                                         |                   |             |
| DLL1           | NONE                                                                         |                   |             |
| IL28A          | NONE                                                                         |                   |             |
| Gene  | Value |
|-------|-------|
| LFGN  | NONE  |
| PIN4  | NONE  |
| ZNF284| NONE  |
References

1. Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, et al. Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res 39: D691-697.
2. Kovacs P, Stumvoll M, Bogardus C, Hanson RL, Baier LJ (2006) A functional Tyr1306Cys variant in LARG is associated with increased insulin action in vivo. Diabetes 55: 1497-1503.
3. Scheuplein F, Rissiek B, Driver JP, Chen YG, Koch-Nolte F, et al. A recombinant heavy chain antibody approach blocks ART2 mediated deletion of an iNKT cell population that upon activation inhibits autoimmune diabetes. J Autoimmun 34: 145-154.
4. Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson A, Manolescu A, et al. (2006) Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet 38: 320-323.
5. Rulifson IC, Karnik SK, Heiser PW, ten Berge D, Chen H, et al. (2007) Wnt signaling regulates pancreatic beta cell proliferation. Proc Natl Acad Sci U S A 104: 6247-6252.
6. Inada A, Yamada Y, Someya Y, Kubota A, Yasuda K, et al. (1998) Transcriptional repressors are increased in pancreatic islets of type 2 diabetic rats. Biochem Biophys Res Commun 253: 712-718.
7. Farook VS, Hanson RL, Wolford JK, Bogardus C, Prochazka M (2002) Molecular analysis of KCNJ10 on 1q as a candidate gene for Type 2 diabetes in Pima Indians. Diabetes 51: 3342-3346.
8. Gloyn AL, Pearson ER, Antcliff JF, Proks P, Bruining GJ, et al. (2004) Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. N Engl J Med 350: 1838-1849.
9. Gloyn AL, Weedon MN, Owen KR, Turner MJ, Knight BA, et al. (2003) Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. Diabetes 52: 568-572.
10. Meissner T, Beinbrech B, Mayatepek E (1999) Congenital hyperinsulinism: molecular basis of a heterogeneous disease. Hum Mutat 13: 351-361.
11. Nestorowicz A, Inagaki N, Gono T, Schoor KP, Wilson BA, et al. (1997) A nonsense mutation in the inward rectifier potassium channel gene, Kir6.2, is associated with familial hyperinsulinism. Diabetes 46: 1743-1748.
12. Oetjen E, Blume R, Cierny I, Schlag C, Kutschenko A, et al. (2007) Inhibition of MafA transcriptional activity and human insulin gene transcription by interleukin-1beta and mitogen-activated protein kinase kinase kinase in pancreatic islet beta cells. Diabetologia 50: 1678-1687.
13. Gardiner JV, Bataveljic A, Patel NA, Bewick GA, Roy D, et al. Prokineticin 2 is a hypothalamic neuropeptide that potently inhibits food intake. Diabetes 59: 397-406.
14. Sookoian S, Gianotti TF, Schuman M, Pirola CJ (2009) Gene prioritization based on biological plausibility over genome wide association studies renders new loci associated with type 2 diabetes. Genet Med 11: 338-343.
15. http://www.genome.jp/kegg/pathway/hsa/hsa04930.html Type II Diabetes Mellitus KEGG pathway.
16. Schreyer SA, Chua SC, Jr., LeBoeuf RC (1998) Obesity and diabetes in TNF-alpha receptor- deficient mice. J Clin Invest 102: 402-411.
17. Skurat AV, Dietrich AD, Zhai L, Roach PJ (2002) GNIP, a novel protein that binds and activates glycogenin, the self-glucosylating initiator of glycogen biosynthesis. J Biol Chem 277: 19331-19338.
18. http://www.reactome.org/entitylevelview/PathwayBrowser.html#DB=gk_current&FOCUS_SPECIES_ID=48887&FOCUS_PATHWAY_ID=381119&ID=381038&VID=2833811 T2D sub-pathway in Reactome.
Supplement Table S7. The Empirical SGAN Frequency (p-value) of 21 genes in Figure 5. \(^1\)P-value is defined by the number of occurrence of partnership of protein in 1,000 re-sampling.

| Host Gene Name | SNP rs number | GWAS of Origin | \(P\)-value\(^1\) |
|----------------|---------------|----------------|-----------------|
| ARG1           | rs17184300    | FUSION         | 0.001           |
| ARHGEF12       | rs11217854    | WTCCC          | 0.001           |
| ART1           | rs6578410     | FUSION         | 0.001           |
| AXIN1          | rs7359414     | FUSION         | <0.001          |
| CDC42          | rs2056975     | WTCCC          | <0.001          |
| CDKL3          | rs4958228     | FUSION         | <0.001          |
| CREM           | rs2505639     | WTCCC          | 0.001           |
| DLL1           | rs1033583     | FUSION         | <0.001          |
| IL28A          | rs664893      | WTCCC          | 0.001           |
| KCNJ10         | rs1130183     | FUSION         | <0.001          |
| KCNJ11         | rs5215        | FUSION         | 0.001           |
| KCNJ11         | rs5215        | WTCCC          | 0.001           |
| LFNG           | rs2895        | WTCCC          | <0.001          |
| MAP3K1         | rs726501      | FUSION         | <0.001          |
| PIN4           | rs6525591     | FUSION         | 0.001           |
| PROK2          | rs3796224     | WTCCC          | <0.001          |
| SNAP29         | rs165598      | FUSION         | 0.001           |
| TACR3          | rs17304065    | WTCCC          | <0.001          |
| TRAP1          | rs17136481    | WTCCC          | <0.001          |
| TRIM7          | rs254456      | FUSION         | <0.001          |
| ZBTB17         | rs10927875    | FUSION         | 0.001           |
| ZNF284         | rs4803674     | WTCCC          | 0.001           |
Supplement Table S8. Edgetic P-value of interactions and Evidence. The edgetic P-value was calculated with a set of genes from both WTCCC-ranked (600 host genes cutoff) and FUSION-ranked (600 host genes cutoff). PIN-Ranked genes are the top prioritized T2D genes of Figure 5 (Panel C).

We report the STRING evidence score (varies from 0 to 999; no evidence to high evidence) for the following types of STRING evidence: cooccurrence, experimental, database, textmining, and their combined scores.

No evidence in STRING of Neighborhood, STRING-annotated Fusion, and Coexpression for 8 interactions. No additional evidences from BIOGRID, REACTOME, MINT, and HRPD for 8 interactions.

| PIN-Ranked Gene 1 | PIN-Ranked Gene 2 | Edgetic P-value | Cooccurence | Experimental | Database | Text mining | Combined score | STRING Version |
|-------------------|-------------------|----------------|--------------|--------------|----------|-------------|----------------|----------------|
| CDC42             | ARHGEF12          | 0.060          |              |              | 967      | 967         | 8.2            |                |
| CDC42             | MAP3K1            | 0.140          |              |              | 760      | 760         | 6.3            |                |
| CDC42             | MAP3K1            | 0.140          |              |              | 956      | 956         | 8.2            |                |
| CDC42             | PIN4              | 0.142          |              |              | 160      | 160         | 6.3            |                |
| CDC42             | ZNF284            | 0.001          | 202          |              |          | 202         | 6.3            |                |
| MAP3K1            | AXIN1             | 0.029          |              | 994          |          | 994         | 8.2            |                |
| LFING             | DLL1              | 0.004          | 875          |              |          | 875         | 8.2            |                |
| DLL1              | KCNJ10            | 0.013          |              | 899          |          | 899         | 8.2            |                |
| KCNJ10            | KCNJ11            | 0.003          | 189          |              |          | 189         | 8.2            |                |
**Supplement Table S9 Statistical Evidence and Protein Interaction.** Evidence that PIN-Prioritized Gene Interact Directly with Known T2D Gene Edgetic $P$-value is a statistical likelihood for the direct interaction of 21 PIN-prioritized genes with 17 known T2D genes under control by the empirical distribution from 10,000 permutation re-sampling. $^1$Edgetic p-value. No evidence in STRING from Neighborhood, Coocurrence, and coexpression. The STRING scores vary in a range from 0 (no evidence) to 999 (high evidence). However STRING documentation does not provide methods from which these scores are derived.

| PIN-prioritized Gene (1st interactor of GS) | Known T2D Gene | Edgetic $P$-value$^1$ | Fusion | Experimental Database | Text mining | Combined score | Version | Other Evidence |
|-------------------------------------------|----------------|-----------------------|--------|-----------------------|-------------|----------------|---------|----------------|
| LFNG NOTCH2                               | 0.0036         | 761 (627)             | 800 (800) | 542                  | 978 (992) | 6.3 (8.2) | Yes   | Yes | |
| DLL1 NOTCH2                               | 0.0182         | 15                    | 942 (900) | 900                  | 995        | 8.2           | Yes   | Yes | Yes |
| ZBTB17 PPARG                              | 0.0276         |                      | 523 (523) | 523                  | 6.3        | Yes           |       |     |     |
| KCNJ10 KCNQ1                              | 0.0310         |                      | 612 (612) | 612                  | 6.3        | Yes           |       |     |     |
| KCNJ10 PPARG                              | 0.0538         |                      | 915 (915) | 915                  | 8.2        | Yes           |       |     |     |
| MAP3K1 PPARG                              | 0.1009         |                      | 156 (156) | 156                  | 6.3        | Yes           |       |     |     |
| Acronym/Term                        | Definition                                                                                                                                                                                                                                                                                                                                 |
|-----------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Betweenness (network metric)      | betweenness is a network metric calculated using an established algorithm ([http://www.gersteinlab.org/proj/bottleneck/](http://www.gersteinlab.org/proj/bottleneck/)). It corresponds to the number of times a node (protein) acts as a bridge along the shortest path between two other nodes. High betweenness of a protein is called bottleneckness, in other words, a protein required as a gatekeeper for a lot of second degree interactions. |
| Bottleneck (Bottleneck protein)   | bottlenecks as genes for which the corresponding proteins are ranked among the top 20% according to the betweenness metric calculated in the PIN.                                                                                                                                                                                                 |
| Centrality property of a network  | Network measures such as hub or bottleneckness. A protein of high centrality is directly (hubness) or indirectly (bottleneckness) required for many protein interactions.                                                                                                             |
| Complex Disease                   | Polygenic disease of complex inheritance patterns. In contrast to single-gene / Mendelian diseases with straightforward autosomal or recessive inheritance patterns.                                                                                                                          |
| eQTL                              | expression Quantitative Trait Locus                                                                                                                                                                                                                                                                                                      |
| Edge, edgetic (network representation) | In this manuscript, relationship between two proteins.                                                                                                                                                                                                                                                                                 |
| Empirical SPAN Frequency          | The statistical likelihood of the observed host gene connectivity in the SPAN analysis; the likelihood of randomly finding the number of interactions identified by SPAN analysis for a protein derived from GWAS identified SNPs                                                                                                                                 |
| Enrichment statistic (genomic)    | Measure of the excess overlap of molecules between two sets of molecules (e.g. genes or proteins). This measure is comparable to a contingency table. Thus odds ratio, chi-square statistics, hypergeometric distribution and the Fisher Exact Tests are alternate approaches utilized to establish its statistical significance. |
| FET                               | Fisher's Exact Test                                                                                                                                                                                                                                                                                                                      |
| FUSION                            | Finland - United States Investigation of NIDDM Genetics. The abbreviation NIDDM is used in the manuscript in the context of to the FUSION dataset name. However, T2D and NIDDM are interchangeable in this manuscript – with the preferred term being T2D.                                                                |
| GO                                | Gene Ontology                                                                                                                                                                                                                                                                                                                          |
| GWAS                              | Genome-Wide Association Study                                                                                                                                                                                                                                                                                                             |
| GWAS-ranked SNP & GWAS-ranked host gene | SNPs or their corresponding host genes ranked by a SNP’s P-value in the original GWAS                                                                                                                                                                                                                                                    |
| Host gene                         | The host genes of intragenic SNPs were defined by genomic boundaries extending from 200 kb upstream (5’ side) to 0.5kb downstream (3’ side) of the gene.                                                                                                                                                                  |
| Host gene cutoff                  | GWAS-ranked host genes prioritized above the input cutoff for SPAN network analyses                                                                                                                                                                                                                                                      |
| Hub, hubness (Hub protein)        | hubness of an intragenic SNP is defined using the connectivity of its host gene (gene for which the corresponding protein is in the top 20% when ranked by node degree)                                                                                                                                                          |
| Intragenic SNP                    | SNP located in a host gene.                                                                                                                                                                                                                                                                                                              |
| KEGG                              | Kyoto Encyclopedia of Genes and Genomes                                                                                                                                                                                                                                                                                                   |
| IBDGNC                            | Inflammatory Bowel Disease Genetics Consortium                                                                                                                                                                                                                                                                                           |
| Acronym/Term | Definition |
|--------------|------------|
| MAF          | Minor Allelic Frequency |
| Network modeling (protein-protein interaction network models) | Computational modeling over PINs using centrality or other metrics. |
| NHGRI GWAS Catalog | A collection of the trait associated SNPs from published GWAS which seek to determine the genetic variants associated with complexly inherited traits, thus this catalog exclusively contains SNP-trait associations for complex traits or disorders |
| NIDDM | Non-Insulin Dependent Diabetes Mellitus. The abbreviation NIDDM is used in the manuscript in the context of to the FUSION dataset name. However, T2D and NIDDM are interchangeable in this manuscript – with the preferred term being T2D. |
| Node (network representation) | In this manuscript, nodes are proteins of the protein interaction networks. |
| Node degree | The network metric: the count of first interactions to a node. In this manuscript: count of direct protein interactors among the prioritized host genes of SNPs (the SNPs that are intragenic are translated to genes, which have a corresponding protein in the PIN from which the node degree is calculated). |
| Odds ratio of network model | Statistical quantity of the re-capitulation of known Type 2 Diabetes genes in the network model (methods) |
| OMIM | Online Mendelian Inheritance in Man |
| Optimal network model | A network model containing the highest number of true and significant signals buried within a large set of SNPs |
| PIN | Protein Interaction Network |
| Reprioritized SNP (SPAN-reprioritized SNP) | SNP with a new prioritization originally ranked by a GWAS according to SPAN network analysis |
| SNP | Single Nucleotide Polymorphism |
| SPAN | Single Protein Analyses in a Network |
| Topological centrality (in PIN) | See centrality |
| Trait, phenotypic trait | Normal or abnormal inheritable phenotypic character (e.g. blue eyes or adult onset diabetes) |
| Trait-associated SNP | SNP confirmed in at least two independent and well-powered GWAS to be associated to a trait |
| T2D | Type 2 Diabetes Mellitus, the abbreviation NIDDM is also used in the manuscript due to the naming of the FUSION dataset. T2D and NIDDM are interchangeable in this manuscript – with the preferred term T2D. |
| WTCCC | Wellcome Trust Case Control Consortium |
**Supplement Methods**

**Supplement details on protein interaction datasets.** In this study, we included only protein interactions that were derived from *Homo sapiens*. To ensure the independence of the network, protein interactions derived only from text mining were removed from STRING version 8.2 since its publication historically followed the publication of the GWAS of interest (WTCCC, Fusion and IBDGC). Text mining results were included in STRING version 6.3 whose publication date historically preceded that of WTCCC and FUSION, but followed IBDGC by four months. Furthermore, publications citing IBDGC from its online publication date to the release of STRING version 6.3 (October 2006-January 2007) were examined to determine if they contained information that would be included via text mining. Papers were examined to determine if they a) cited IBDGC, b) contained the names of two proteins, and c) contained this information in a PubMed abstract (STRING’s criteria). Since, none of these publications met the criteria, the text mining results included in STRING vers. 6.3 would not be supported by IBDGC results. Duplicate protein interaction entries and symmetrical relationships were refined so that only one interaction was included.

**Supplement details of the Empirical control for protein network model (related to Figure 1).** To conduct an empirical control for our T2D network analysis we created 1,000 resampled empirical SNP lists and derived their corresponding list of host genes. Intragenic SNPs were resampled a thousand times without replacement (1,000 bootstraps) within the total of 187,842 from WTCCC, 137,248 from FUSION, and 134,247 from IBDGC. The observed sets of host genes at different cutoffs serve as distinct inputs for network modeling as do those from bootstrapping. Therefore, to avoid the bias of PIN degree of genes corresponding to SNPs, the intragenic SNPs are sampled until they generate the same number of host genes present in the PIN as the ones observed in the study at each rank cutoff. Since a fixed list of SNPs may yield a slightly variable number of host genes, due to the fact that multiple SNPs that belong to one gene may be sampled, the sampling cutoffs for sets of host genes associated with GWAS-prioritized SNPs could be fixed at either 1) the number of SNPs or 2) the number of host genes. By design, we opted for the latter to obtain better network model controls with fixed host gene list sizes.

**Supplement details of the odds ratios of SPAN network models (related to Figure 4).** In each GWAS, one network model is produced for each host gene cutoff. Within this network and at
this host gene cutoff, the selected GWAS-ranked host genes are reprioritized according to the likelihood of observing their connectivity by bootstrapping, that we term the empirical SPAN frequency. The subset of host genes and their associated original GWAS-ranked SNPs within each network model are then further refined and divided into smaller sets at different empirical SPAN frequencies (≤ 0.1%, 0.5%, 1%, 3%, 5%, 7%, and 9%). Within these associated original GWAS-ranked SNPs reprioritized by SPAN at each host gene cutoff and empirical SPAN frequency, reprioritized SNPs are considered true positives when found among the gold standard SNPs derived from the NHGRI and false positive if not. Accordingly, gold standard genes not among reprioritized GWAS SNPs are considered false negatives. The FET was used to calculate each network model’s odds ratio and the \( P \)-values of each set according to two previously described network model parameters: host gene cutoff and empirical SPAN frequency. The background used for these calculations contains all the intragenic SNPs for which a gene was found in the protein interaction dataset.

**Supplement details of Single Protein Analysis of Networks (SPAN).** In order to properly control for the connectivity of each protein in our real network, we performed 1,000 bootstraps in which the connections for each protein were randomized simultaneously while the node degree was kept constant. In other words, each hub protein is properly controlled, as it remains a hub in each permutation. For each bootstrap, we selected a set of host genes translated from randomized SNPs from WTCCC, FUSION or IBDGC to generate each network using a node randomization approach. In our network, proteins are considered nodes and interactions between proteins are edges. Since biological networks are scale-free rather than random, node randomization can create conservative “permuted nodes” as controls, from which we can derive an empirical distribution of interactions between a subset of proteins. 1,000 bootstrapped gene sets were generated from the original background SNPs consisting of real datasets from each respective GWAS. The real dataset consists of host genes selected using the GWAS-ranked SNP with the best (lowest) \( P \)-value among all SNPs annotated to the gene.

Each of these host genes was translated to its corresponding protein identifier in the network. For the real dataset, each protein was then mapped to each of its interacting proteins according to existing pairs of protein interactions in the PIN yielding an Observed number of distinct Protein Interactions (Observed count of PI). Thereafter, the same procedure was applied to the 1,000
empirical gene sets yielding control counts of distinct protein interactions for each of the genes translated from the randomized SNPs (Control count of PI).

For each protein, a $P$-value was assigned by measuring the frequency at which the “Observed count of PI” of that protein occurred in the empirical distribution’s “Control counts of PI” (1000 total) for each specific protein. Each protein thus is assigned its own individual $P$-value and was subsequently ranked according to this $P$-value. At each $P$-value cutoff, a certain number of proteins were prioritized. Consequently, a FDR of the prioritized proteins was calculated by dividing the median number of proteins prioritized at that cutoff in the empirical distributions of the randomized PINs divided by the observed number of prioritized proteins in the real PIN. We refer to this approach as single protein analysis in the network (SPAN) since each gene’s partnerships are randomized simultaneously, allowing for a proper control of each individual gene’s connections, or node degree, in the network.

**Supplement details of the optimal SPAN model of T2D and its evaluation (related to Figure 4B, C): Calculation of edgetic P-values in SPAN.** First degree protein interactors, are shown biologically to be more functionally similar than non-interactors, and are used to identify putative T2D intragenic SNPs. The frequency at which the genes of the Optimal SPAN Model of T2D were found as first interactors to these independent known T2D genes was calculated using 10,000 permutation resamplings of the network. As we previously described, the number of interactors of each specific gene remains constant in each resampling providing a conservative empirical distribution well-controlled for the connectivity of each gene (node degree). In the 10,000 permutation re-samplings of all protein-protein pair, we count how many times each protein-protein pair appears in each generated random network. Then for each observed pair of proteins, we obtain a $p$-value for the likelihood of each protein-protein pair’s occurrence by dividing the number of times the two proteins are paired in all permuted networks by 10,000. For the 10,000 permutation re-samplings, we sorted the p-value in ascendant then aggregate counts. So for each observed protein-protein pairs, we get edgetic FDR equal median aggregate count divided by observed aggregate count.
Supplement details of the GWAS SNPs reprioritized by protein interaction models are more likely to be validated in ulterior studies. The optimal network model is selected based on the model’s odds ratio of identifying gold standard SNPs discovered in ulterior, independent GWAS annotated in the NHGRI catalog. To ensure the independence of the gold standard, SNPs derived from ulterior re-analyses or meta-analyses of the SPAN modeled GWAS are excluded. Based on our empirical distributions, we identified the host gene cutoff and empirical SPAN frequency parameters for selecting the optimal network model that contains the highest number of true and significant signals buried within a large set of SNPs. Specifically, after we established our SPAN models according to the size of the GWAS-ranked host gene set and its frequency of protein interaction found via SPAN analysis, we evaluated them according to our gold standard. We calculated the odds ratio of reprioritized host genes’ corresponding SNPs in the SPAN model with empirical SPAN frequencies ≤ 0.1%, 0.5%, and 1%. The odds ratio of finding gold standard genes in each unmodified set of GWAS-ranked SNPs was also calculated as a control for network models since they denote the maximum number of gold standard SNPs that could be identified by the intragenic SNPs genotyped in a given set.

Supplement details of calculation of correlation between centrality of recombination rates of genes. Human gene annotations (locations in genomic sequences and gene symbol ids) were download from the UCSC Genome browser (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/refGene.txt.gz) on June 15, 2011 and gene recombination rates were downloaded from HapMap (http://hapmap.ncbi.nlm.nih.gov/downloads/recombination/latest/rates/) on August 31, 2011. The start and end positions of each gene were obtained by merging the positions of all overlapping alternative spicing copies. Only the first copy of a gene was taken for genes with multiple segregated copies. The left and right nearest markers of the gene (nearest to the two ends of the gene) were identified from gene recombination rate data from HapMap. The recombination rate of each gene was calculated using the recombination rate at the regions between the two nearest markers of the gene, which is quantified by the distance in the genetic map (in centimorgans (cM)) divided by the genomic distance (in units of a million base pairs) where both quantities were extracted from HapMap. Hubness is the number of interacting proteins in the protein interaction network. The bottleneckness was calculated by publically available tools
(http://www.gersteinlab.org/proj/bottleneck; more in the end of Method Section of the manuscript). Out of the 21450 genes that were used in the analysis 12968 genes had overlapped with 14025 genes in protein interaction networks.