Supplementary Methods 2

Supplementary Methods 2

Notes on the Network Analysis in this Letter

We report here a structured knowledge base network analysis of the human transcriptional response to systemic inflammation. This analytic approach is based on a comprehensive examination of experimental microarray data by the systematic utilization of previously published scientific findings extracted in the Ingenuity Pathways Knowledge Base (KB).

The KB is constructed through the efforts of Ph.D.-level scientists who have read the abstracts of every paper in the Ingenuity KB. These scientists manually extracted 87% of the findings in the KB from the full text of >200,000 articles, including the abstract, text, tables, and figures. The remaining 13% were extracted by first natural language processing and subsequently, were manually verified against the abstracts from which the findings were derived. As of January 2005, the KB includes information of more than 9,800 human (including the ~9,500 Reviewed and Validated Human RefSeqs, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=gene), 7,900 mouse, and 5,000 rat genes.

All information in the KB is curated from the published literature. Every gene interaction in our report is supported by evidence extracted from the underlying publications, structured using an ontology, and stored in the KB. The literature findings are available online in the Ingenuity Pathways Analysis (IPA) program. Importantly, each finding is categorized as a study of either human, mouse or rat genes (based on Entrez Gene annotation, http://www.ncbi.nlm.nih.gov/entrez/), and supported by a literature citation and a link to the PUBMED abstract.

The quality of the experimental work is not interpreted by the curator(s), as the goal of the knowledge base is to represent the published scientific findings, demonstrated by the authors, and not to re-interpret or bias these findings in any way. Scientists using the KB can view and evaluate the conflicting findings and the experimental techniques used by parsing the original publications. The false-positive rate for the interactome primarily depends on the false-positive rates of the direct experimental assays used to demonstrate the interactions. Content from assays with substantial error rates (e.g. single-pass yeast two-hybrid screens) is excluded a priori from the KB to ensure high accuracy. Internal sampling and quality control[1] has estimated the false-positive rate due to error in curation process to be less than 1%.

Network analysis was performed on molecular relationships involving 8,000 human orthologs (between human, mouse and rat, as defined by Homologene, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=homologene). While some of these relationships have not been directly tested in the human interactome, they occur between

orthologous mammalian genes that are likely to be isofunctional and used as approximating human biology[2,3].

Specialized cells such as blood leukocytes are expected to include cell-type specific gene interactions. At the genome-wide scale, only a very small percentage of gene functions and gene interactions have been studied experimentally in a specific cell type. As an approximation, the relationships in the interactome are based on evidence from different cell types in human, rat and mouse; therefore they should be taken as representing 'potential' interactions in a specific cell type. Novel networks inferred on the basis of merging the interactome with cell-specific expression data lead to testable hypotheses that wait to be further verified experimentally.

Readers can use the network visualization, search and browsing tools of IPA available online to directly evaluate our analytic approach as well as the KB. The system is available to the public through trial and subscription access. Readers can request access to the IPA by requesting a trial account at http://www.ingenuity.com/trial.

Readers can independently reproduce our results by uploading our gene dataset (the file "List of Significant Genes and Verification.txt" available at the supplementary website at http://www.gluegrant.org), and regenerating the network analysis results themselves in IPA. Please note that since new content is continuously being added to the KB on an ongoing basis, this option may be of particular interest to readers who are interested in re-analyzing our data in the context of the most recent knowledge-base content. In addition, readers can evaluate our approach against other datasets by using IPA to upload and analyze new high throughput datasets.

For additional help on reproducing our results please contact the authors. We appreciate your questions and comments.

References

1. U.S. patent 6,741,986, "Method and system for performing information extraction and quality control for a knowledgebase"
2. Waterston, R.H. *et al*. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520-562 (2002).
3. Gibbs, R.A. *et al*. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493-521 (2004)

1. Examining KB content for RELA (Suppl Figure 1)

Supplementary Figure 1 describes the network neighborhood surrounding a single gene, RELA. To reproduce this network in IPA, please perform the following steps identified below. The network neighborhood is derived from findings extracted from the public literature and stored in the KB. Every gene interaction is supported by evidence from the underlying papers.

Gene	Neighboring Genes	Direct Interactions	Findings	Papers
RELA	150	619	7,118	847

Please note that the above table in the report is based on KB content statistics from September, 2004, and new content is being added to the KB on an ongoing basis (the following examples are taken from January, 2005).

To view the textual web page view of KB information on a gene and its related gene products and complexes:

1. Log into the IPA system.
2. From the "Click on the Name of an Analyzed Dataset" (for example, "Systemic Inflammation | LPS-treated human leukocytes") to go to the "Analysis Summary" page.
3. From the "Analysis Summary" page, click the "Search" tab.
4 From the "Search" tab, select "Ingenuity Pathways Knowledge Base", type the short name or symbol of your gene (e.g., "RELA" or "XXXX"), and click Search.

5. The "Search Results" displays genes that match your criteria (use "*" for wildcard searches). There are three values under "Exp Val":
    a. the first is for coloring the node of the gene based on the bin number (see Figure 1). Genes in bins 0-4 are colored blue, and in bins 5-9 are colored red.
    b. the second shows whether the gene is verified in the verification experiment. A label of "1.000" indicates that the gene is also identified as significant in the verification experiment, and a label of "-1.000" indicates that it is not significant in the second experiment (FDR <0.01))
    c. the third shows the bin number of the gene (0-9)

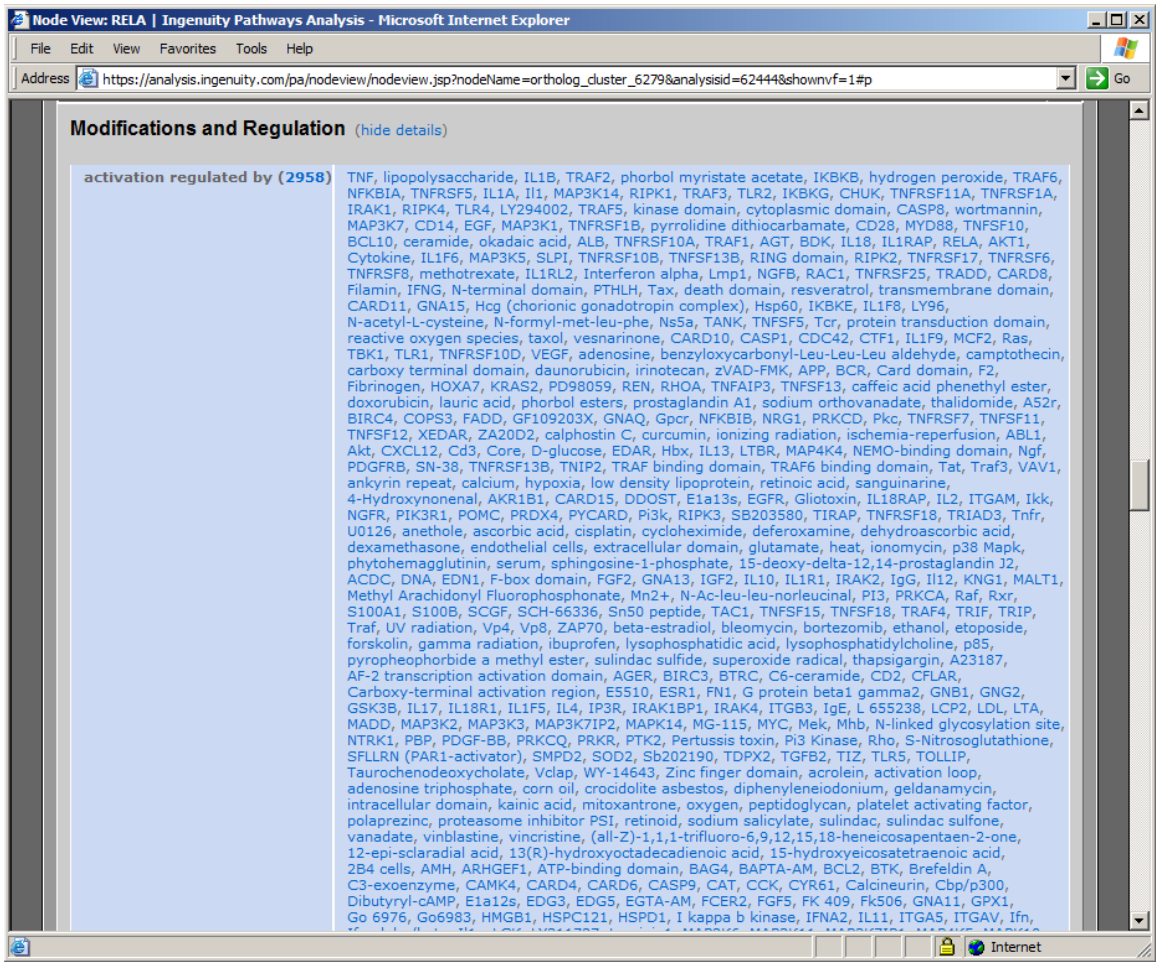
Gene (show synonyms)	Identifier	Exp Val			Networks	Description	Family	Location	Drugs
	LocusLink	Other	Other	Other					
RELA	5970	↑7.000	1.000	7.000	3, 30, 41, 46, 56, 62	v-rel reticuloendotheliosis viral oncogene homolog A, nuclear factor of kappa light polypeptide gene enhancer in B-cells 3, p65 (avian)	transcription regulator	Nucleus	--

6. From the "Search Results", click the "Gene" name to go to the Node View page for that Gene.
    a. This page includes the number of findings for that gene. Click on "Show Details" at the bottom of the page to view a detailed breakdown of findings about this gene.
    b. Click on the Human / Mouse / Rat tabs to see species-specific information,

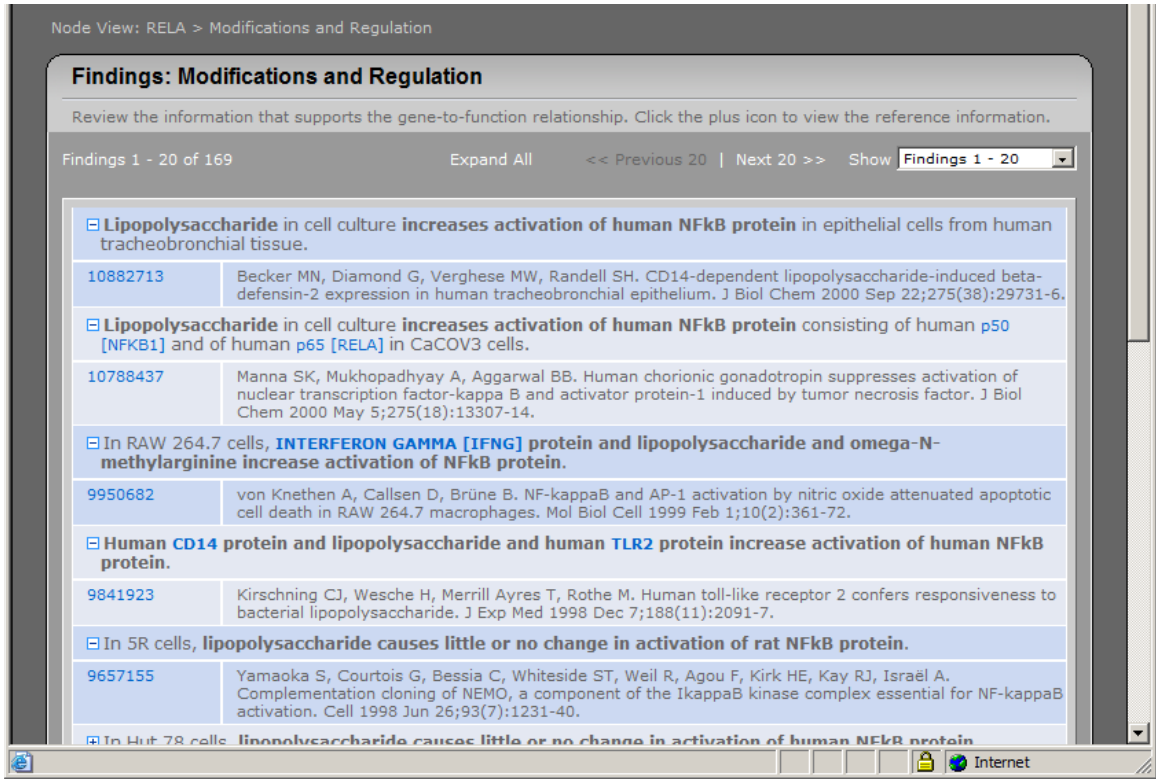
c. Click on Neighborhood Explorer to view the graphical network of molecular interactions for this gene.

Below is an example of a Node View for RELA, which contains categorized literature findings about RELA as well as its related gene products and complexes (e.g. NFkB) (9155 findings by January, 2005):

To view the findings, click on “show all 9155 categorized literature findings”. “Node View” findings are categorized into sections based on topic. For example, the “Activation regulated by” subsection in the “Modifications and Regulation” section of the RELA Node View (shown below), which contains 2958 findings about compounds and genes that regulate the activation of RELA:

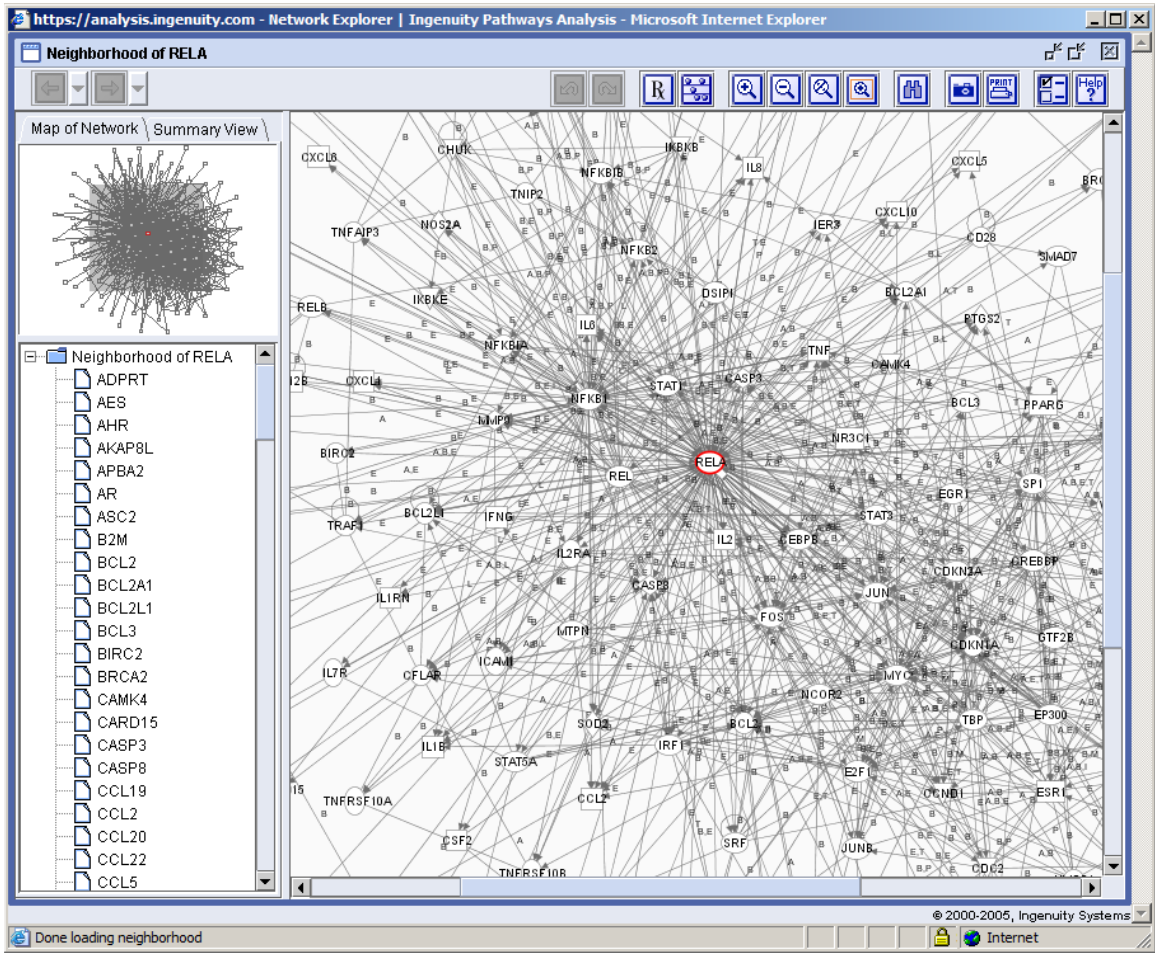


Every entry in a Node View page is supported by published literature findings. For example, clicking on the first link in the above screenshot shows the 169 literature findings in KB that describe RELA activation by lipopolysaccharide. Every finding is supported by a literature reference and a link to the PUBMED abstract:



2. Examining the Network Neighborhood for RELA (Suppl Figure 1)

Supplementary Figure 1 describes the network neighborhood surrounding a single gene, RELA. The Network Neighborhood is a graphical display of the subset of KB relationships for a gene that describes molecular interactions between that gene and its neighboring genes or gene products.



To reproduce this figure and view the graphical nature of the molecular network encompassing RELA, perform the following steps (1 – 5 are the same as above):

1. Log in to IPA.
2. From the “Click on the Name of an Analyzed Dataset” (for example, “Systemic Inflammation | LPS-treated human leukocytes”) to go to the “Analysis Summary” page.
3. From the “Analysis Summary” page, click the “Search” tab.
4. From the “Search” tab, select “Ingenuity Pathways Knowledge Base”, type the short name or symbol of your gene (e.g., “RELA”), and click Search.
5. From the “Search” results, click the “Gene Name” to go to the “Node View” page for RELA.

6. From the “Node View” page, click the “Neighborhood Explorer” link at the top of “Node View” to view the Network Neighborhood Explorer for the selected gene (RELA). The “Network Neighborhood” includes all the network edges connecting a gene to its interacting neighbors.
7. From the Network Neighborhood Explorer, with the selected node in red in the center:
    a. Zoom in / out by clicking on the magnifying glass icons at the top.
    b. The “Map of Network” tab on the left shows a list of all nodes in this network.
    c. Click on a gene to see the “Node Summary” panel in the top-left, including a summary of that gene and its functions.
    d. Click on an edge to see the “Edge Summary” panel in the bottom-right, a summary of the functional and physical relationships between those genes.
    e. From the “Node Summary” panel, click on “Neighborhood Explorer” link to show a new network around that node.

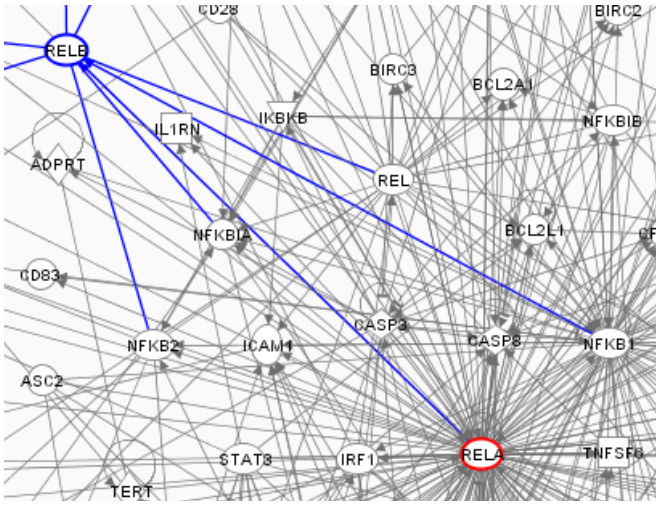
Using this approach, it is possible to browse and examine the entire KB network: start from one gene, examine its neighborhood, click on a neighboring gene of interest to display its “Node Summary”, and click the “Neighborhood Explorer” to jump to a new network neighborhood.

3. Examining the Innate Immunity Network (Figure 2)

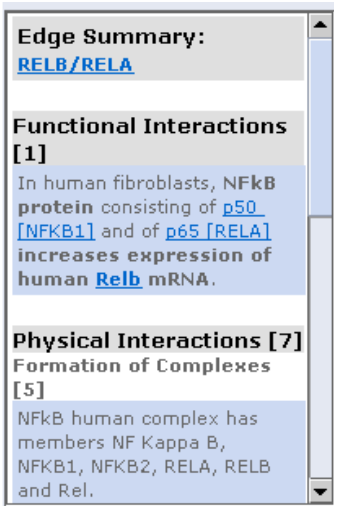
To examine the literature evidence underlying the relationships in Figure 2 (or any other network), please select a pair of genes of interest, use the steps outlined above to view the "Network Neighborhood" for the first gene, locate the second gene in the diagram, and click the edge / link that connects them to see the supporting findings.

As an example, RELA and RELB (near the center of Figure 2) are selected as a pair of genes of interest. Shown below are the steps to examine the literature evidence underlying the relationship between these two genes. Please note that using the same approach, one can browse and examine the entire Innate Immunity Network from Figure 2.

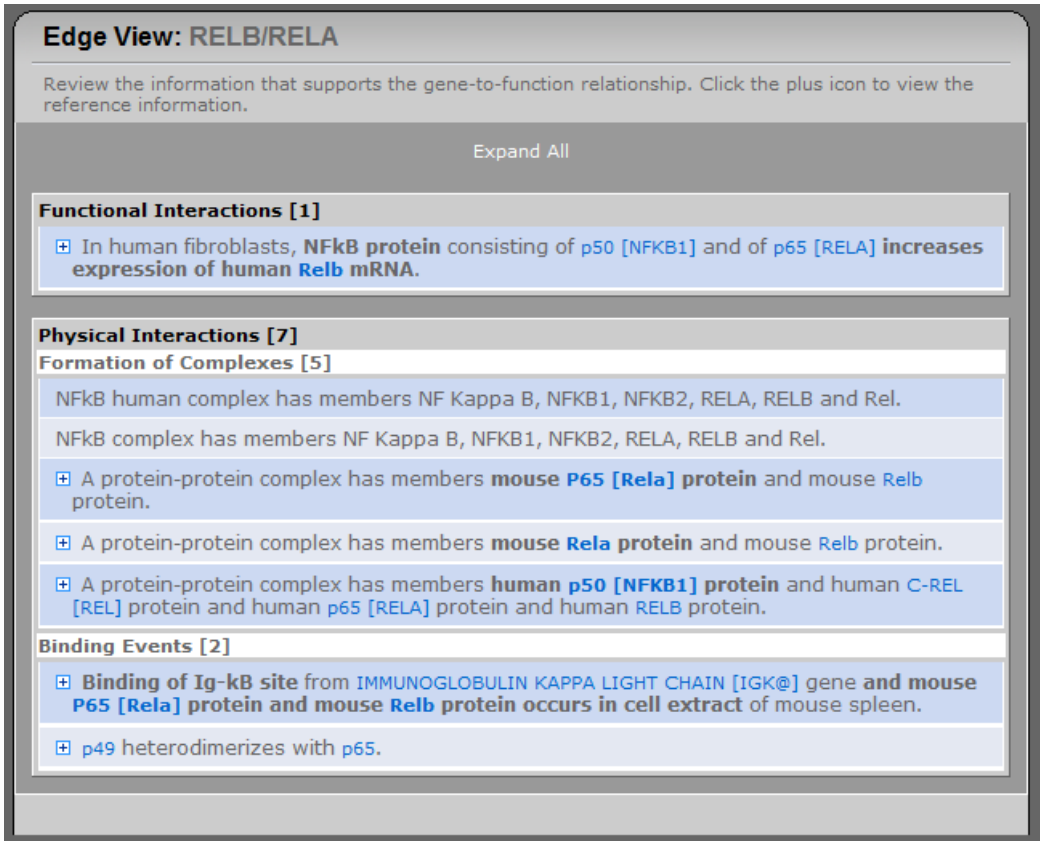
1. Log in to IPA.
2. Click on the "Systemic Inflammation | LPS-treated human leukocytes" analysis results to go to the "Analysis Summary" page.
3. From the "Analysis Summary" page, click the "Search" tab.
4. From the "Search" tab, select "Ingenuity Pathways Knowledge Base", type the short name or symbol of your gene (e.g., "RELA"), and click "Search".
5. From the "Search" results, click the "RELA" link to go to the "Node View" for that gene:
6. From the "Node View" page for "RELA", click "Neighborhood Explorer" at the top of the page to view the "Network Neighborhood Explorer" for your gene.
7. From the "Network Neighborhood Explorer", with RELA in red in the center, locate RELB in the diagram and click on it. Alternately, select "RELB" in the Map of Network list on the left. RELB is highlighted in blue, as are the edges to its neighbors, including RELA:



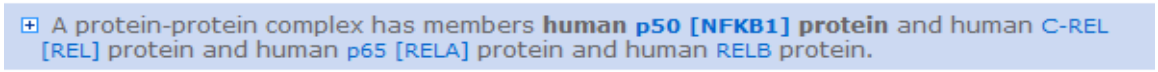
8. Click on the edge connecting RELA to RELB to display the "RELB/RELA" Edge Summary. Edge Summary provides a brief overview of the relationship between two gene products. Clicking on any edge opens the edge summary window on the left side of the screen:



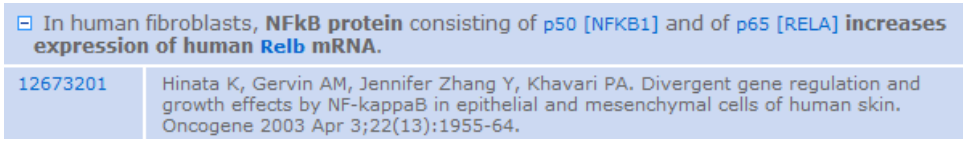
9. To view the evidence supporting an edge relationship between two genes, click on the GeneA/GeneB link (e.g. the "RELB/RELA" link) at the top of "Edge Summary" to reach "Edge View". "Edge View" provides a more detailed overview of the relationships between two gene products than "Edge Summary". All of the KB findings along with their references supporting a given edge are listed in "Edge View":



Importantly, every finding is categorized and labeled as a study of either human, mouse or rat genes.



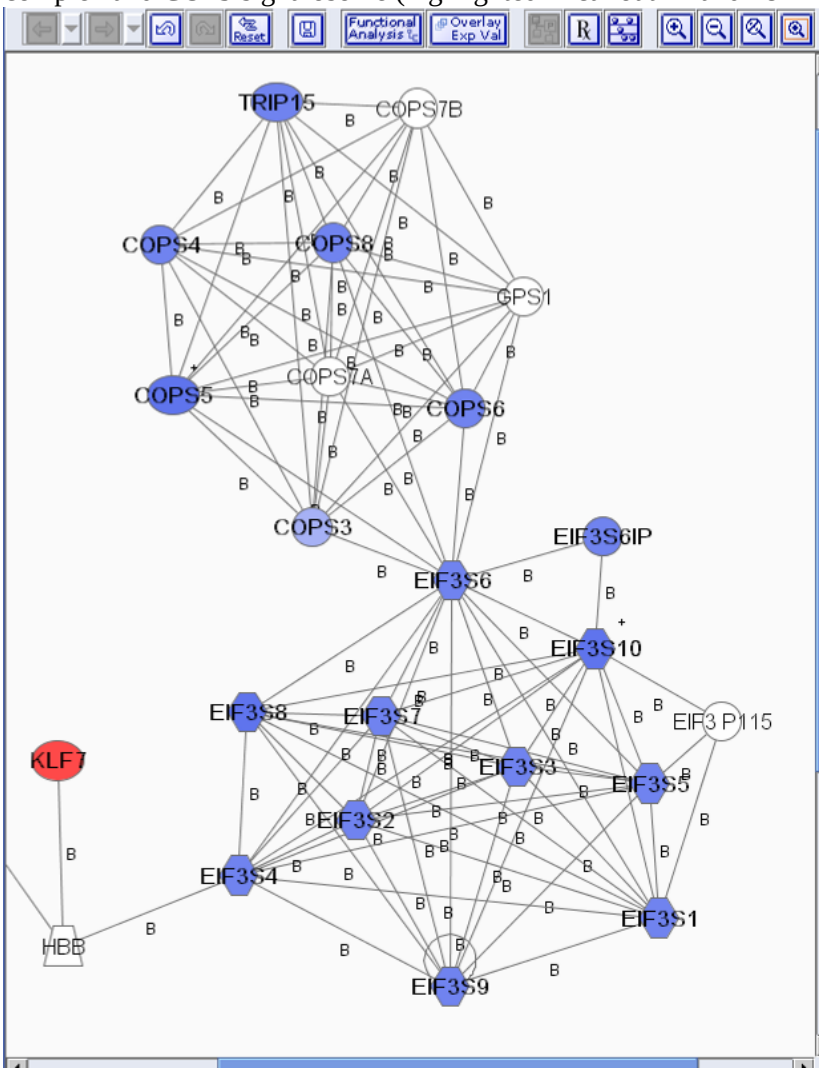
10. Click on the [+] symbol next to each finding to see the literature reference, and on the Pubmed ID next to each literature reference to view the MEDLINE abstract:



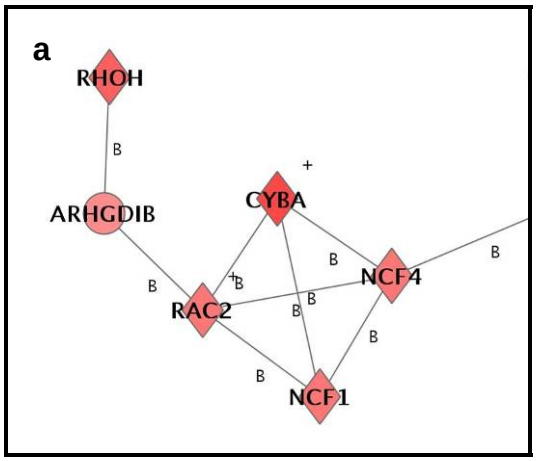
Using this approach, it is possible to browse and examine the entire Innate Immunity Network from Figure 2: find a relationship of interest, view the neighborhood for one of the genes, locate the second gene in the network, and click in on its edge to begin "drilling down" to view the supporting evidence establishing a relationship between the two genes of interest..

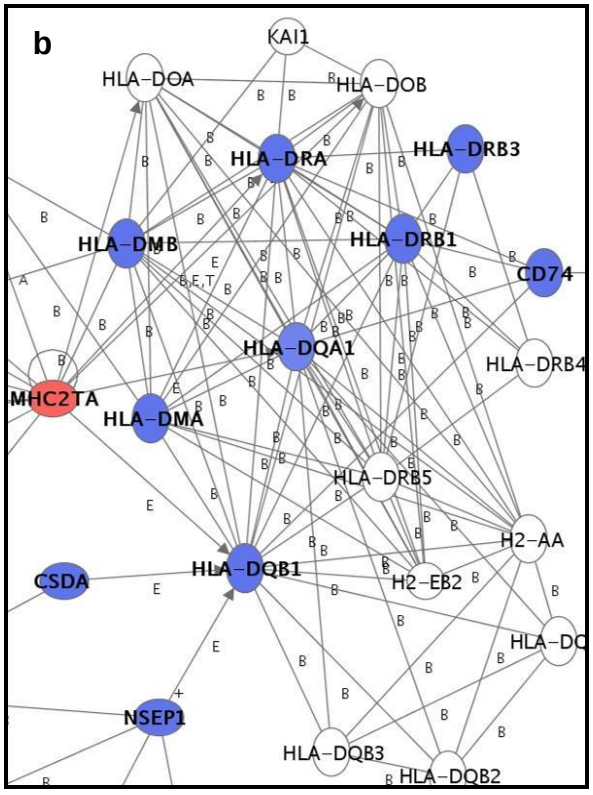
4. Examining the Global Network Analysis Results (Figure 3)

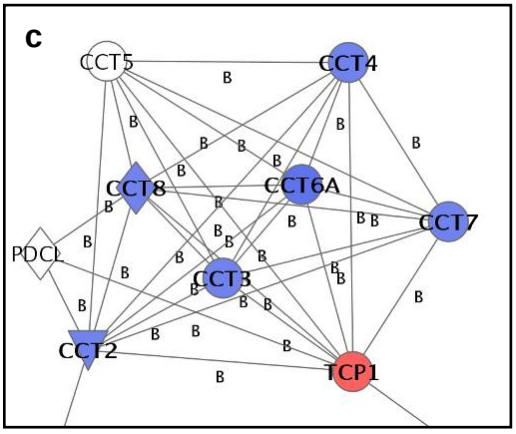
The global network in Figure 3 was constructed by combining the top scoring pathway subnetworks generated by this analysis (see report for details). Below is an example of a generated subnetwork (Network #26), which corresponds to the elongation initiation factor 3 complex and COP9 signalosome (highlighted in callout #7 and #8 in Figure 3):

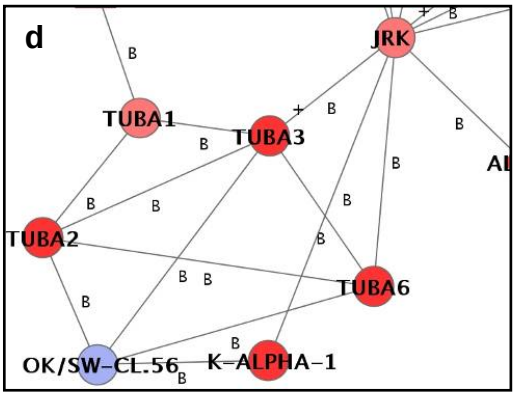


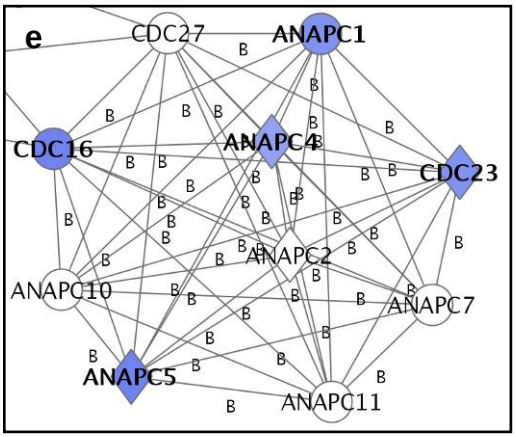
Other examples of the regions discussed in the manuscript include: **a.** the superoxide-producing phagocyte NADPH-oxidase, **b.** members of the MHC II complex were suppressed, **c.** the TCP1 ring complex, **d.** tubulin-A genes, **e.** subunits of the anaphase-promoting complex, **f.** integrin chains.

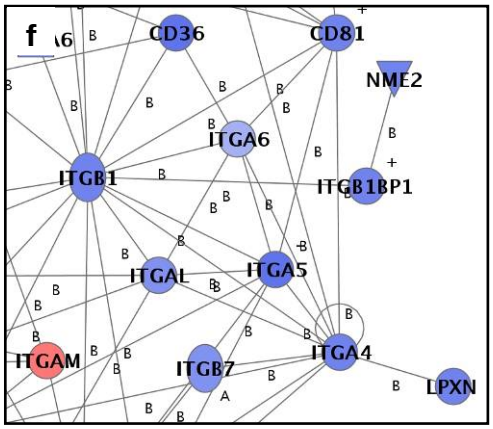

a
RHOH
ARHGDIB
CYBA
NCF4
RAC2
NCF1


b
KAI1
HLA-DOA
HLA-DOB
HLA-DRA
HLA-DRB3
HLA-DMB
HLA-DRB1
CD74
HLA-DQA1
HLA-DRB4
MHC2TA
HLA-DMA
HLA-DRB5
H2-AA
CSDA
HLA-DQB1
H2-EB2
NSEP1
HLA-DQB3
HLA-DQB2


c
CCT5
CCT4
CCT8
CCT6A
CCT7
PDCL
CCT3
CCT2
TCP1

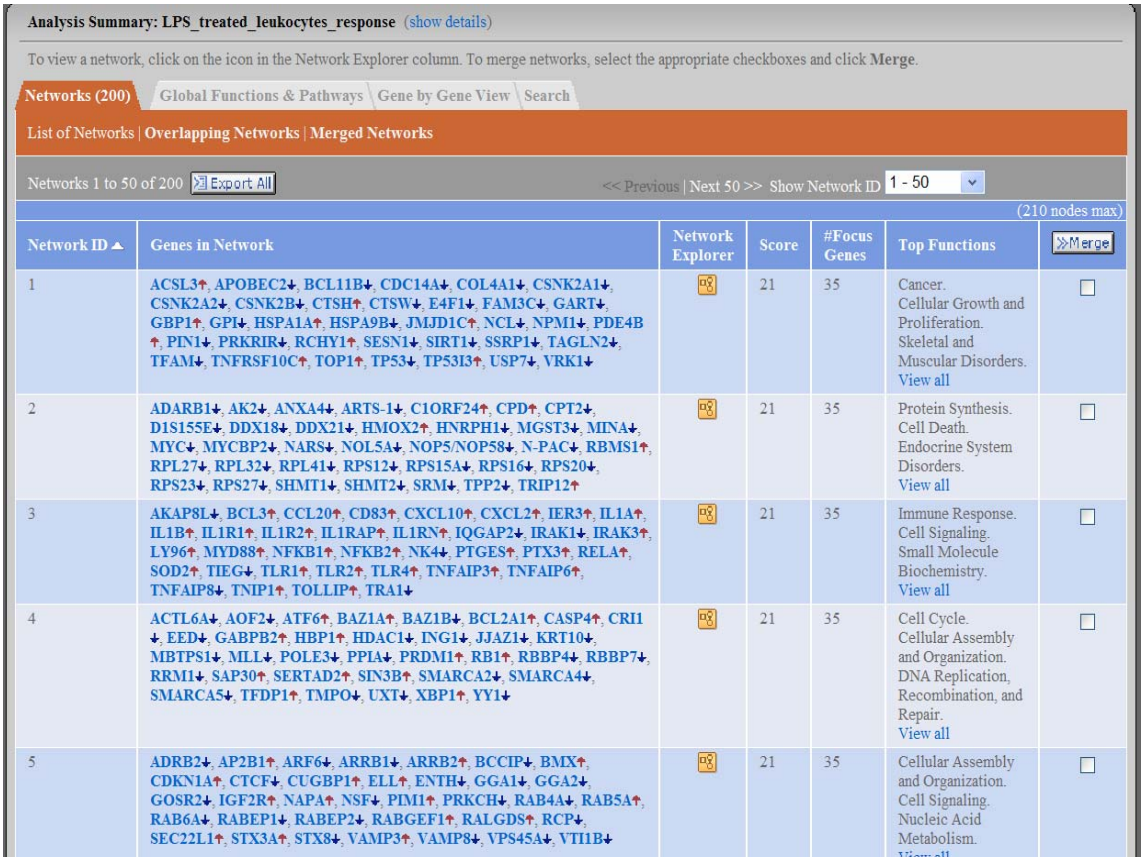

d
JRK
TUBA1
TUBA3
TUBA2
TUBA6
OK/SW-CL.56
K-ALPHA-1


e
CDC27
ANAPC1
CDC16
ANAPC4
CDC23
ANAPC2
ANAPC10
ANAPC7
ANAPC5
ANAPC11


f
CD36
CD81
NME2
ITGA6
ITGB1
ITGB1BP1
ITGAL
ITGA5
ITGAM
ITGB7
ITGA4
LPXN

To examine the full set of pathway networks generated by this analysis:

1. Log in to IPA.
2. Click on the "Systemic Inflammation | LPS-treated human leukocytes" analysis results to go to the "Analysis Summary" page. The "Analysis Summary" page displays a list of all subnetworks generated by the analysis.
3. From the "Analysis Summary", select a network and click on the gold icon in the "Network Explorer" column to view the network diagram.
4. The "Network Explorer" visualization applet will display the network and provide access to the underlying KB findings in the same way as described in the previous section.



Analysis Summary: LPS_treated_leukocytes_response (show details)

To view a network, click on the icon in the Network Explorer column. To merge networks, select the appropriate checkboxes and click **Merge**.

Networks (200) | Global Functions & Pathways | Gene by Gene View | Search

List of Networks | **Overlapping Networks** | **Merged Networks**

Networks 1 to 50 of 200 Export All << Previous | Next 50 >> Show Network ID 1 - 50

(210 nodes max)

Network ID ▲	Genes in Network	Network Explorer	Score	#Focus Genes	Top Functions	>>Merge
1	ACSL3↑, APOBEC2↓, BCL11B↓, CDC14A↓, COL4A1↓, CSNK2A1↓, CSNK2A2↓, CSNK2B↓, CTSH↑, CTSW↓, E4F1↓, FAM3C↓, GART↓, GBP1↑, GPI↓, HSPA1A↑, HSPA9B↓, JMJD1C↑, NCL↓, NPM1↓, PDE4B↑, PIN1↓, PRKRIR↓, RCHY1↑, SESN1↓, SIRT1↓, SSRP1↓, TAGLN2↓, TFAM↓, TNFRSF10C↑, TOP1↑, TP53↓, TP53I3↑, USP7↓, VRK1↓		21	35	Cancer. Cellular Growth and Proliferation. Skeletal and Muscular Disorders. View all	
2	ADARB1↓, AK2↓, ANXA4↓, ARTS-1↓, C1ORF24↑, CPD↑, CPT2↓, D1S155E↓, DDX18↓, DDX21↓, HMOX2↑, HNRPH1↓, MGST3↓, MINA↓, MYC↓, MYCBP2↓, NARS↓, NOL5A↓, NOP5/NOP58↓, N-PAC↓, RBMS1↑, RPL27↓, RPL32↓, RPL41↓, RPS12↓, RPS15A↓, RPS16↓, RPS20↓, RPS23↓, RPS27↓, SHMT1↓, SHMT2↓, SRM↓, TPP2↓, TRIP12↑		21	35	Protein Synthesis. Cell Death. Endocrine System Disorders. View all	
3	AKAP8L↓, BCL3↑, CCL20↑, CD83↑, CXCL10↑, CXCL2↑, IER3↑, IL1A↑, IL1B↑, IL1R1↑, IL1R2↑, IL1RAP↑, IL1RN↑, IQGAP2↓, IRAK1↓, IRAK3↑, LY96↑, MYD88↑, NFKB1↑, NFKB2↑, NK4↓, PTGES↑, PTX3↑, RELA↑, SOD2↑, TIEG↓, TLR1↑, TLR2↑, TLR4↑, TNFAIP3↑, TNFAIP6↑, TNFAIP8↓, TNIP1↑, TOLLIP↑, TRA1↓		21	35	Immune Response. Cell Signaling. Small Molecule Biochemistry. View all	
4	ACTL6A↓, AOF2↓, ATF6↑, BAZ1A↑, BAZ1B↓, BCL2A1↑, CASP4↑, CRI1↓, EED↓, GABPB2↑, HBP1↑, HDAC1↓, ING1↓, JJAZ1↓, KRT10↓, MBTPS1↓, MLL↓, POLE3↓, PPIA↓, PRDM1↑, RB1↑, RBBP4↓, RBBP7↓, RRM1↓, SAP30↑, SERTAD2↑, SIN3B↑, SMARCA2↓, SMARCA4↓, SMARCA5↓, TFDP1↑, TMPO↓, UXT↓, XBP1↑, YY1↓		21	35	Cell Cycle. Cellular Assembly and Organization. DNA Replication, Recombination, and Repair. View all	
5	ADRB2↓, AP2B1↑, ARF6↓, ARRB1↓, ARRB2↑, BCCIP↓, BMX↑, CDKN1A↑, CTCF↓, CUGBP1↑, ELL↑, ENTH↓, GGA1↓, GGA2↓, GOSR2↓, IGF2R↑, NAPA↑, NSF↓, PIM1↑, PRKCH↓, RAB4A↓, RAB5A↑, RAB6A↓, RABEP1↓, RABEP2↓, RABGEF1↑, RALGDS↑, RCP↓, SEC22L1↑, STX3A↑, STX8↓, VAMP3↑, VAMP8↓, VPS45A↓, VTI1B↓		21	35	Cellular Assembly and Organization. Cell Signaling. Nucleic Acid Metabolism.	

In addition to viewing individual networks, several networks can be merged together by selecting the checkboxes in the rightmost column and clicking the "Merge" button.

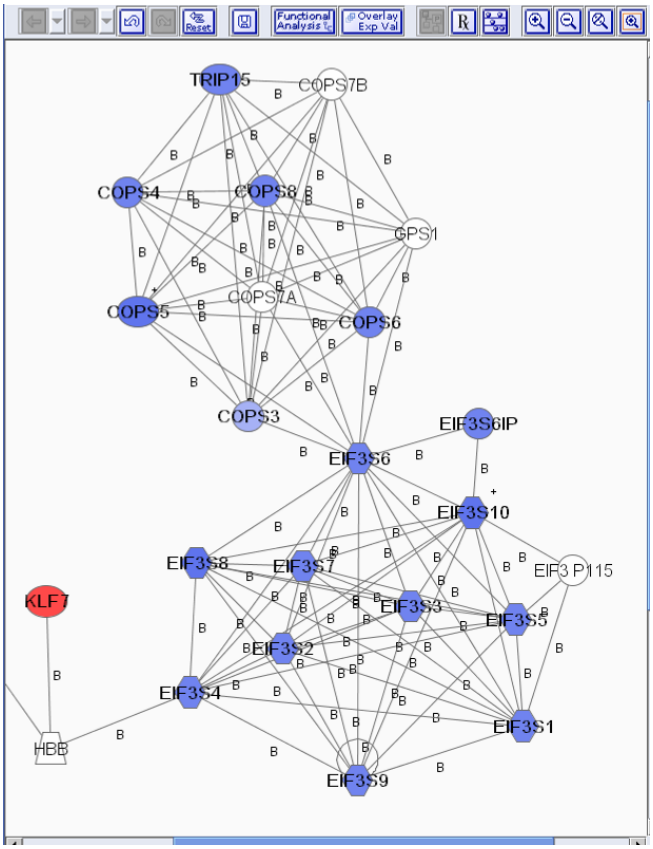
To find a sub network that contains a particular gene, search for genes within the analysis results. For example, to locate the network that contains the elongation initiation factor 3 complex genes (e.g. EIF3S5) displayed in callout #7 in Figure 3, perform the following steps:

1. From the "Analysis Summary" page, click the "Search" tab.
2. From the "Search" tab, select "Networks", and type the short name or symbol of the gene (e.g., "EIF3S5"), and click "Search":

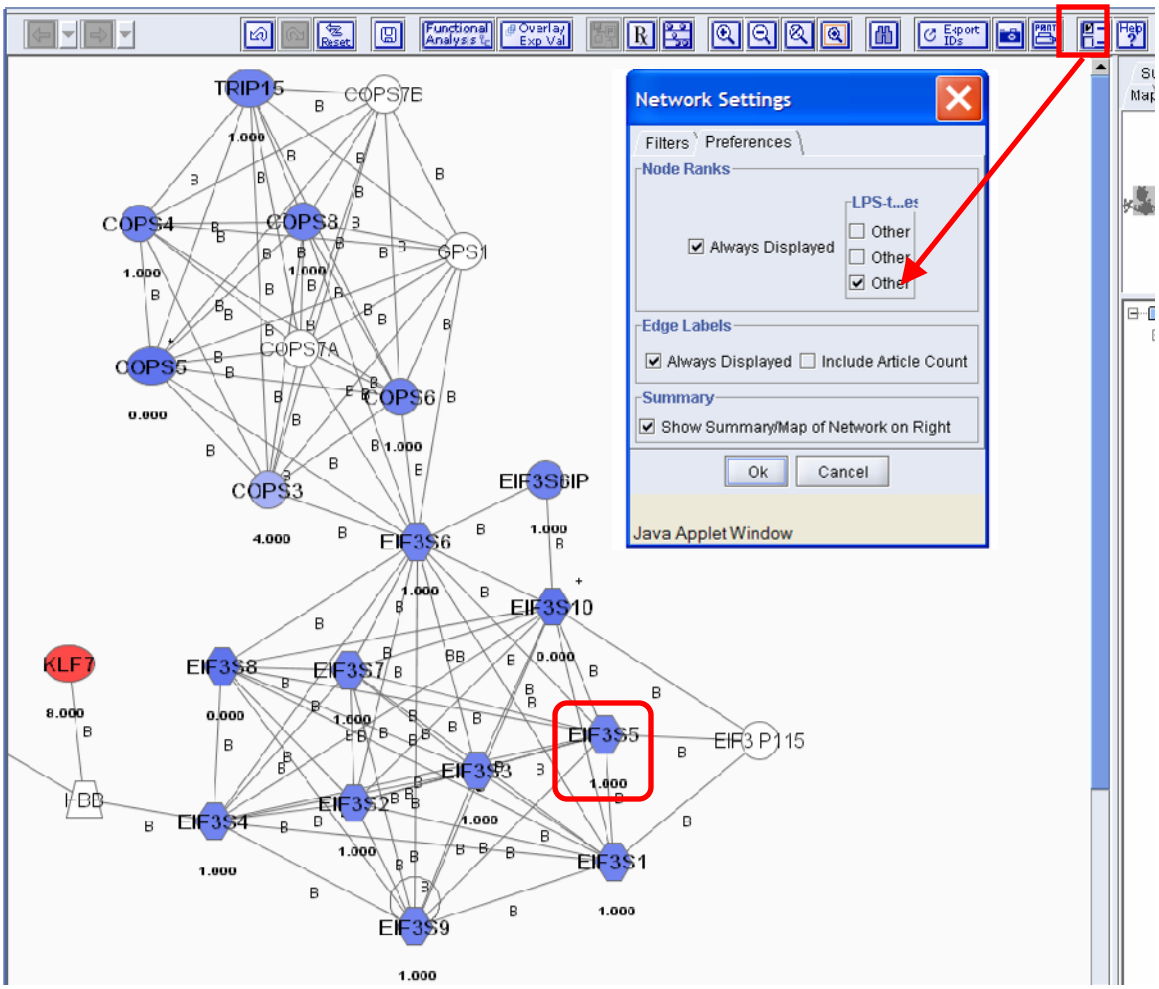
3. The "Search" Results display networks that match the criteria (use "*" for wildcard searches). It is also possible to search for multiple gene names, which will identify networks that contain all of the genes:

Network ID	Genes in Network	Network Explorer	Score	#Focus Genes	Top Functions
26	**COPS3↓**, **COPS4↓**, **COPS5↓**, **COPS6↓**, COPS7A, COPS7B, **COPS8↓**, EIF3 P115, **EIF3S1↓**, **EIF3S10↓**, **EIF3S2↓**, **EIF3S3↓**, **EIF3S4↓**, **EIF3S5↓**, **EIF3S6↓**, **EIF3S6IP↓**, **EIF3S7↓**, **EIF3S8↓**, **EIF3S9↓**, GPS1, HBB, **KLF7↑**, **NONO↓**, NR5A1, SF1, **SFPQ↓**, **TOP1↑**, **TRIP15↓**		7	21	Protein Synthesis. Embryonic Development. Tissue Development. View all

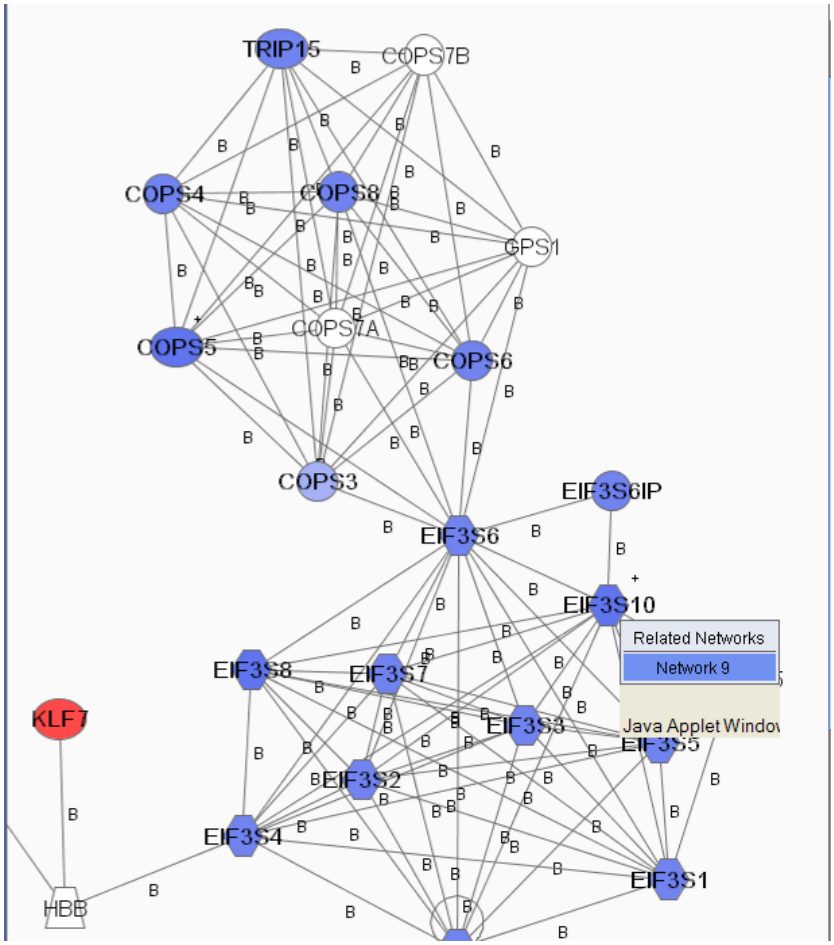
4. From the "Search" results, click the gold "Network Explorer" icon to display the network diagram for this network. The "Network Explorer" visualization applet will display the network and provide access to the underlying knowledge-base findings in the same way as described in the previous sections:



5. To display the Bin numbers of the genes, click on the “Network Settings” icon (second icon from the right side) to bring up the “Network Settings” window. In the “Preference” menu, select the check box of the third “Other”, which brings up the Bin number below each gene identified as significant.



6. Please note that the “+” sign above EIF3S10 and COPS5, indicating that these genes are shared between more than one of the sub-networks. For example, right click on node “EIF3S10”, to reveal the Related Networks, “Network 9”. Click on “Network 9” to display the associated sub network.



Using this approach, one can browse through all the associated sub-networks and examine the entire Global Network in Figure 3A: find a gene of interest, locate a sub network that contains this gene, “drill down” to view the underlying supporting findings in the knowledge-base in the same way as described in the previous sections, and also examine the associated sub-networks of the gene.

5. Examining the Verification Results (Figure 3)

To examine whether a gene (or a sub-network region) of the global network analysis results shown in Figure 3 is confirmed in the follow-up verification experiment (see Supplementary Information S1):

1. Locate a sub network that contains this gene as described in the previous sections
2. Click on the “Network Settings” icon to bring up the settings window.
3. Select the “Preference” tab, and select the check box of second “Other”, which brings up the result of comparison labeled below each gene.

A label of “1.000” indicates that the gene is also identified as significant in the verification experiment, and a label of “-1.000” indicates that it is not significant in the second experiment (FDR <0.01).