

pSTIING: a ‘systems’ approach towards integrating signalling pathways, interaction and transcriptional regulatory networks in inflammation and cancer

Aylwin Ng, Borisas Bursteinas, Qiong Gao, Ewan Mollison and Marketa Zvelebil*

Bioinformatics and Systems Biology Group, Ludwig Institute for Cancer Research, UCL Branch, London W1W 7BS, UK

Received August 15, 2005; Revised and Accepted October 4, 2005

ABSTRACT

pSTIING (<http://pstiing.licr.org>) is a new publicly accessible web-based application and knowledgebase featuring 65 228 distinct molecular associations (comprising protein–protein, protein–lipid, protein–small molecule interactions and transcriptional regulatory associations), ligand–receptor–cell type information and signal transduction modules. It has a particular major focus on regulatory networks relevant to chronic inflammation, cell migration and cancer. The web application and interface provide graphical representations of networks allowing users to combine and extend transcriptional regulatory and signalling modules, infer molecular interactions across species and explore networks via protein domains/motifs, gene ontology annotations and human diseases. pSTIING also supports the direct cross-correlation of experimental results with interaction information in the knowledgebase via the CLADIST tool associated with pSTIING, which currently analyses and clusters gene expression, proteomic and phenotypic datasets. This allows the contextual projection of co-expression patterns onto prior network information, facilitating the identification of functional modules in physiologically relevant systems.

INTRODUCTION

The advent of high-throughput genomic and proteomic approaches coupled with annotation efforts has generated a comprehensive ‘parts list’ of biomolecules encoded by the genome, resulting in the establishment of a number of publicly available repositories and databases, each focussing on a certain specialized aspect of biological data collection, ranging

from gene information [e.g. UniGene (1)], protein entries [e.g. UniProt (2)] and structural domain information [e.g. Pfam (3), InterPro (4) and CATH (5)] to molecular interactions [e.g. MINT (6) and BIND (7)] and pathways [e.g. BioCarta (www.biocarta.com) and KEGG (8)].

The extraction of meaningful information in the context of physiological systems requires characterization of not only the function of individual genes or proteins but also the physical interactions and functional associations in which they participate. To this end, recent genome-wide interactome projects involving two-hybrid (9), co-immunoprecipitation, protein chip analyses (10), affinity purification (11,12) and mass spectrometry (13) have produced a wealth of protein–protein interaction data.

Cellular and physiological processes are complex systems (14) modulated by signals from the extracellular environment and coordinated by the interplay of intracellular regulatory networks assembled into functional modules (15) and pathways. Regulatory networks comprise molecular (protein–protein, protein–lipid, protein–small molecule) interactions and transcriptional regulatory associations between transcription factors and the protein products of genes they *trans-activate*.

In order to understand these processes as interconnected and interdependent systems, there is a need to derive new associations from not only molecular interaction data but also integrating this with transcriptional regulatory associations, signal transduction modules and other information with corroborating evidence from multiple diverse sources, which often entails bringing together heterogeneous data. This is necessary to facilitate effective cross-correlation with expression and functional analyses in physiologically relevant investigations.

In light of this, we developed Protein, Signalling, Transcriptional Interactions and Inflammation Networks Gateway (pSTIING), a publicly accessible knowledgebase and web application to facilitate an integrative ‘systems’ approach towards understanding inflammation, cell migration and cancer. The knowledgebase integrates protein–protein,

*To whom correspondence should be addressed. Tel: +44 20 7878 4012; Fax: +44 20 7878 4040; Email: marketa@ludwig.ucl.ac.uk

protein–lipid, protein–small molecule, ligand–receptor interactions, receptor–cell type associations, disease information and transcriptional associations with signal transduction modules. Experimental data from gene expression or proteomic analyses can also be directly cross-correlated with network information in pSTIING via CLADIST, a clustering and data analysis tool associated with pSTIING. pSTIING-CLADIST is accessible at <http://pstiing.licr.org>.

DATA SOURCES AND DATABASE ACCESS

pSTIING contains data manually curated from the literature and publicly available data derived from multiple sources as listed under six broad categories in Table 1. As pSTIING is growing rapidly, we expect the numbers and data sources listed in Table 1 to change very quickly. Because our current strategy for building pSTIING is to adopt a selective approach rather than taking the ‘all-inclusive’ route towards data integration, we highlight in Table 1 the various subsets of data that were included or not incorporated into pSTIING as it stands currently.

For example, we incorporated 2200 experimentally verified transcriptional associations from TRED that have been ranked highly for promoter and binding quality (16). Lower ranked entries such as from computational predictions or those supported by indirect or unreliable evidence are not incorporated.

To facilitate efficient storage and retrieval of interaction data, we constructed a relational database in an SQL (Oracle) server. The model underlying the database structure at its core consists of modules, interactions and interacting entities. A module (e.g. IL6 signalling module or STAT3 transcriptional module) is a biologically or functionally relevant grouping of interactions or associations. For example, we are building a curated collection of signalling modules by not only providing curated information at the pathway level [as with BioCarta, KEGG (8), STKE (www.stke.org) and AfCS (www.signaling-gateway.org)] but also focussing on individual interactions and associations within signalling modules. The assembly of interactions into functionally relevant signalling modules is an on-going effort involving manual literature curation and also using BioCarta, KEGG, STKE and AfCS as general reference resources.

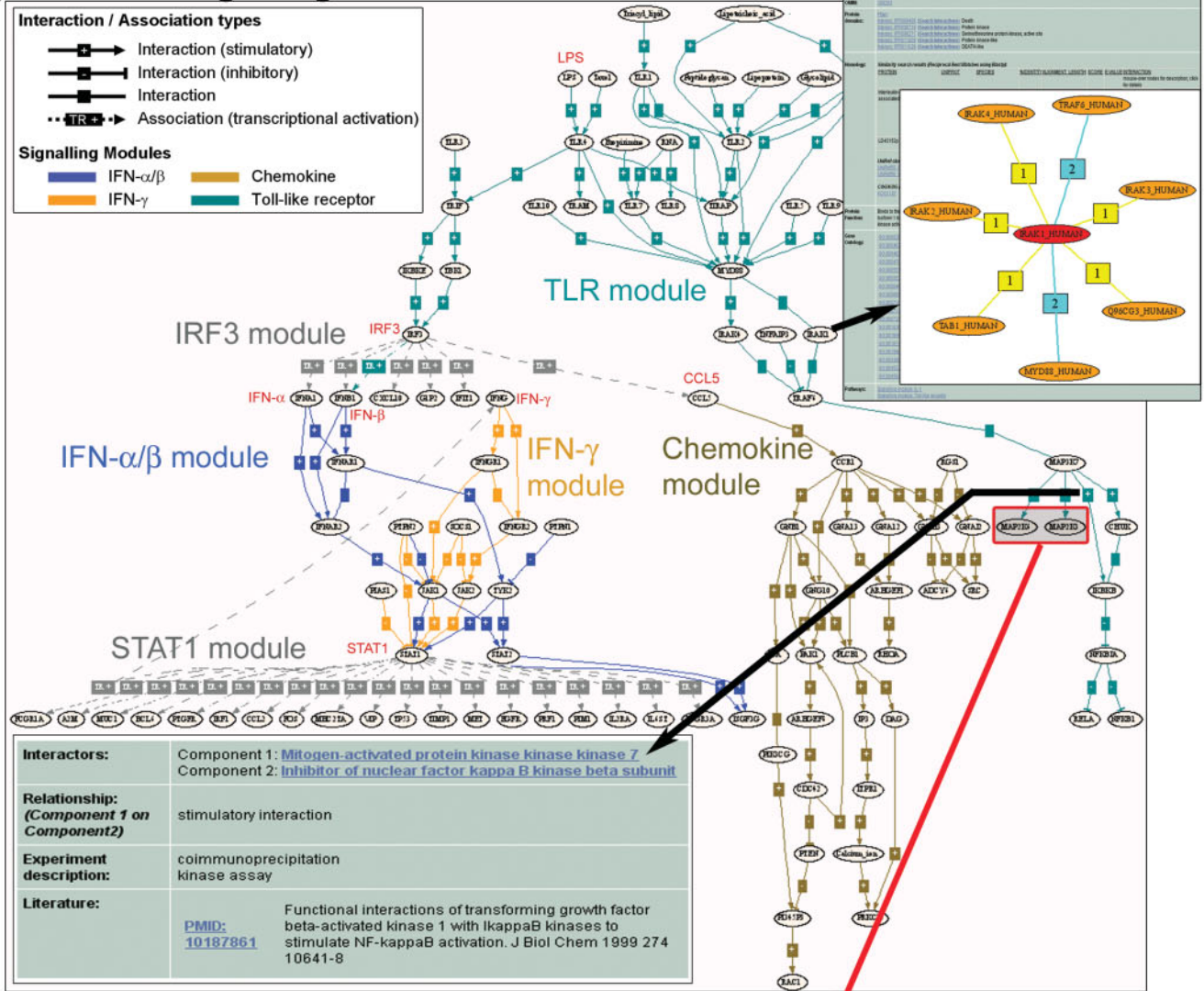
Interactions between interacting entities can be direct (e.g. physical, stimulatory or inhibitory) or indirect (e.g. transcriptional associations). Each interaction has a source (e.g. MINT, HPRD or our in-house curated set) and is supported by experiments which have Pubmed ID links. Currently, most interactions in pSTIING are binary but can also involve more than two interacting entities, which can be any small molecule (e.g. Ca^{2+}) or biomolecules (including proteins and lipids). Small molecules and lipids are linked to KEGG compound identifiers. Proteins are linked to protein-related information

Table 1. Data sources

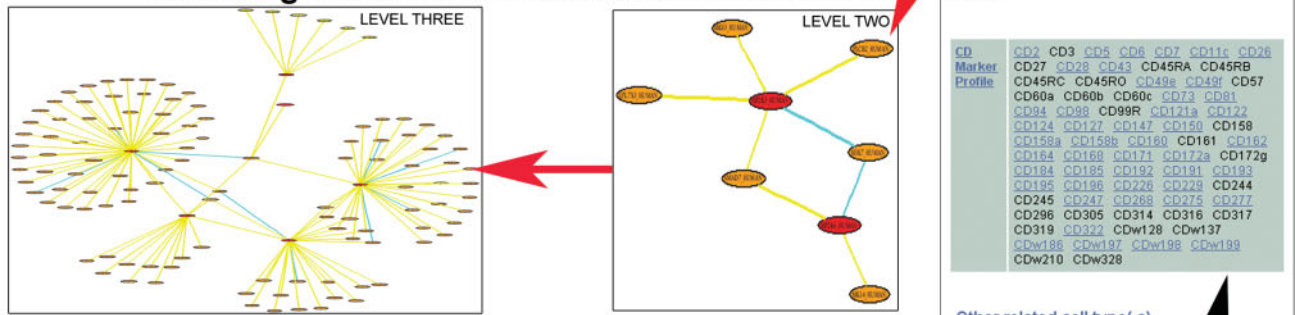
Categories	Data sources	Data incorporated into pSTIING
General protein and gene-related information	UniProt (2)	180 823 entries from key organisms ^a and <i>S.pombe</i> .
	UniRef90 (2)	111 850 entries from key organisms ^a and <i>S.pombe</i> .
	UniRef50 (2)	74 650 entries from key organisms ^a and <i>S.pombe</i> .
	InterPro (4)	7309 entries.
	Entrez Gene (30)	1 280 648 entries.
	Gene/HGND (31)	23 914 entries.
	UniGene (1)	53 256 clusters.
Orthologous groups	GOA (18)	19 141 entries.
	KOG/COG (19)	4852 orthologous groups involving 26 477 proteins from <i>H.sapiens</i> , <i>M.musculus</i> , <i>D.melanogaster</i> , <i>C.elegans</i> . Entries lacking SwissProt/TrEMBL identifiers are excluded.
Human disease Molecular interactions	OMIM (20)	3466 human disease entries.
	MINT (6)	47 403 interactions involving 17 774 interacting entities from key organisms ^a .
	BIND (7)	11 568 interactions involving 6161 interacting proteins from key organisms ^a . Self-interactors and entries lacking SwissProt/TrEMBL identifiers are excluded. Currently, non-protein–protein interactions are not included. Protein–small molecule interactions from BIND will be included soon.
	Reactome (32)	11 452 interactions (‘reaction’ type) involving 520 interacting entities from <i>H.sapiens</i> . Self-interactors and indirect interactions excluded. Macromolecular complexes from Reactome will be included soon.
	HPRD (33)	7011 interactions involving 5516 interacting entities from <i>H.sapiens</i> . Self-interactors or entries lacking SwissProt/TrEMBL identifiers and interactors ambiguously classified as ‘mammalian’ rather than from a clearly defined species are excluded. Macromolecular complexes from HPRD will be included soon.
	In-house curated set	709 protein–protein [including ligand–receptor (34,35)], protein–small molecule, protein–lipid interactions manually curated from the literature.
	MIPS (36)	461 interactions from <i>H.sapiens</i> , <i>R.norvegicus</i> , <i>M.musculus</i> . Self-interactors lacking SwissProt/TrEMBL identifiers and interactions with low-confidence MIPS score are excluded.
Transcriptional and Signalling modules and pathways	In-house curated set	Currently 11 signalling modules (<i>H.sapiens</i>) involving 561 interactions manually curated from the literature. Other reference sources: BioCarta, KEGG, STKE, AfCS.
	TRED (16)	2200 experimentally verified transcriptional associations (<i>H.sapiens</i>) were incorporated, of which 568 have been placed into transcriptional modules in pSTIING. Associations determined by indirect evidence or prediction methods are excluded.
	KEGG (8) BioCarta	21 signalling pathways involving 1441 proteins (<i>H.sapiens</i>). 241 signalling pathways involving 1165 proteins (<i>H.sapiens</i>) from BioCarta (www.biocarta.com).
Receptor, CD marker and Cell type information	In-house curated set	797 CD marker/receptor–cell type associations manually curated from the literature and PathologyOutlines (www.pathologyoutlines.com).
	HCDM/HLDA8	358 CD marker entries from HCDM (www.hlda8.org).

^aKey organisms: *H.sapiens*, *R.norvegicus*, *M.musculus*, *D.melanogaster*, *C.elegans* and *S.cerevisiae*.

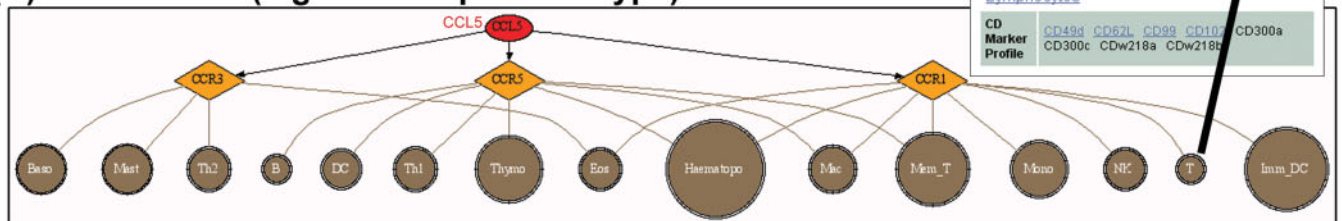
(A) Intracellular signalling modules



Extending networks in the direction of selection



(B) Extracellular (Ligand-Receptor-Cell type) associations



[via UniProt (2) ID], protein domain information [via InterPro (4) accession], protein structure [via PDB (17) ID], gene ontology (GO) [via GOA (18) ID], orthology [via KOG (19) ID], human disease [via OMIM (20) ID] and taxonomy classification [via Taxonomy (1) ID]. Organisms represented in pSTIING include *Homo sapiens* (Taxonomy ID: 9606), *Rattus norvegicus* (10116), *Mus musculus* (10090), *Drosophila melanogaster* (7227), *Caenorhabditis elegans* (6239) and *Saccharomyces cerevisiae* (4932).

The pSTIING web interface provides a number of starting points from which to navigate the database (Supplementary Data 1). Users may perform single or multiple (batch) searches using protein or gene name, synonyms or identifiers (such as SwissProt or TrEMBL accessions) which retrieve information about interacting entities, interactions, transcriptional associations and pathway modules, with supporting experimental evidence and extensive cross-references to external databases (Supplementary Data 2). Users can also query across species, via GO annotations, InterPro domains, motifs, human diseases or using datasets from gene expression or proteomic experiments.

VISUALIZING INTERACTIONS, TRANSCRIPTIONAL ASSOCIATIONS AND EXTENDING PATHWAYS

pSTIING dynamically generates graphical representations of interaction networks and displays maps in both compressed and scalable vector graphics (SVG) formats. With a suitable SVG viewer plug-in for browsers such as Corel SVG or Adobe SVG viewers, users will be able to zoom in and out, or move (pan) a network map around within the viewing screen. Nodes represent interacting components (e.g. proteins, macromolecular assemblies, lipids or small molecules) and edges represent interaction or transcriptional regulatory relationships between components.

In signalling modules, these edges may be directed, denoting stimulatory or inhibitory interactions, or undirected, representing molecular interactions that are neither stimulatory nor inhibitory, or where their stimulatory/inhibitory status is not known. In pSTIING, the representation of a transcriptional association is abstracted to indicate the relationship between two proteins, one of which is a transcription factor that *trans*-activates the expression of the gene encoding the other protein. Such associations are represented by dashed edges. Depending on the respective display mode selected, edges are colour-coded to highlight a specific module or a cellular compartment, within which an interaction occurs, or to indicate the number of documented experimental types supporting an interaction.

Interaction maps can also be progressively expanded outwards to display the next interaction neighbourhood or subsequent levels beyond, thus, making it possible to 'grow'

networks or extend signalling pathway modules in a desired direction. For example, by selecting mitogen-activated protein kinase kinase 3 (MAP2K3) and MAP2K6 from a list of interactors in the Toll-like receptor (TLR) signalling module, the network can be extended to display subsequent interaction levels with MAP2K3 and MAP2K6 at the centre of the interaction network (Figure 1A). By clicking the 'Display Network ('Backbone')' button, it is possible to display only the query nodes and their immediate interaction partners that link query nodes together. This feature is useful for simplifying complex network maps.

LINKING TRANSCRIPTIONAL ASSOCIATIONS TO SIGNALLING MODULES

Regulatory networks comprise both transcriptional regulatory modules and molecular interaction cascades organized within signal transduction modules. Therefore, the incorporation of transcriptional regulatory associations with molecular interaction information is necessary to connect seemingly distinct signalling modules into functional regulatory systems. In pSTIING, multiple signalling and transcriptional regulatory modules can be displayed in combination, facilitating the identification of important 'cross-talk' points (or 'hubs') between modules.

As an example, Figure 1A illustrates this cross-module connectivity between TLR, interferon (IFN) α/β , IFN- γ and chemokine signalling modules participating in an inflammatory response against gram-negative bacteria infection. Bacterial lipopolysaccharide interacts with TLR4 and triggers a cascade of signalling events in the TLR pathway, leading to activation of the transcription factor, IRF3. IRF3 binds to its *cis* element in the promoter of genes encoding type I IFNs (IFN- α and IFN- β) and C-C chemokine, CCL5, inducing their expression. The type I IFN signalling cascade induces the production and secretion of type II IFNs (IFN- γ) via the transcription factor STAT1. IFN- γ triggers a second wave of STAT1 activation which reinforces the first in a positive feed-back loop. Also, CCL5 promiscuously activates three G-protein-coupled receptors CCR1, CCR3 and CCR5 which are expressed on a variety of immunologically relevant cell types, thus transmitting extracellular cues to appropriate effector cells as represented in pSTIING (Figure 1B and Supplementary Data 3). Clicking on a 'cell type' node retrieves information about cell surface (CD or Cluster of Differentiation) markers that phenotypically identify this cell type.

PROJECTING AND MAPPING INTERACTIONS ACROSS SPECIES

The integration of homology information with molecular interaction data allows us to project and map interactions

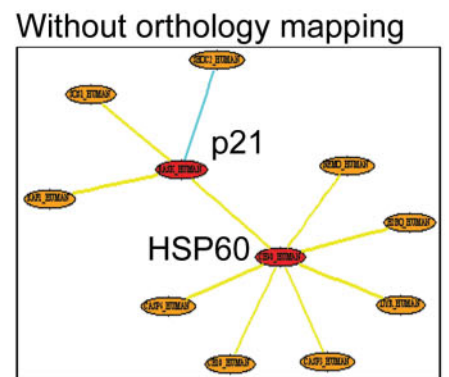
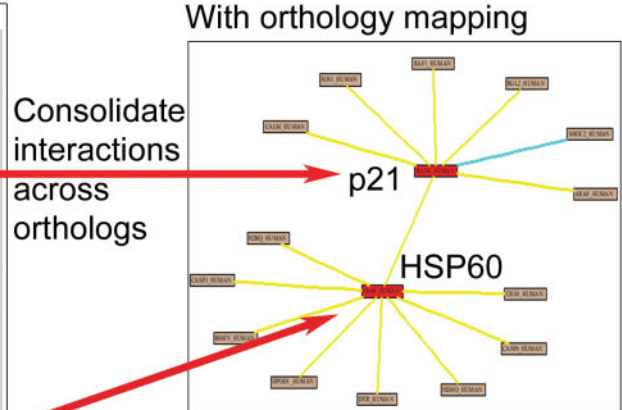
Figure 1. Graphical maps generated in pSTIING showing (A) cross-connectivity between four intracellular signalling modules (TLR, IFN- α/β , IFN- γ and Chemokine) and two transcriptional regulatory modules (STAT1 and IRF3). Interactions are represented by solid edges while transcriptional associations by dashed lines. Each component node (e.g. IRAK1) is linked to more extensive information about the component (including interaction partners, InterPro domains, GO, homologous proteins and cross-links to external databases). Clicking on an interaction node provides interaction details and experimental evidence supporting that interaction. Pathways can be extended in a desired direction by displaying subsequent interaction neighbourhood levels centered on interactors selected by users (e.g. mitogen-activated kinase kinases, MAP2K3 and MAP2K6). (B) Representation of ligand-receptor-cell type associations. The ligand CC-chemokine CCL5 activates a subset of leukocytes and lymphocytes by binding to chemokine receptors CCR1, CCR3 and CCR5 expressed on these cell types. Each ligand and receptor node is linked to detailed information about the respective protein and their interaction partners. Information about each cell type, its related or parental cell types and the cell surface markers (CD antigens) that phenotypically identifies the cell type can be accessed by clicking on the cell type node.

(A) Projecting Interactions Across Species

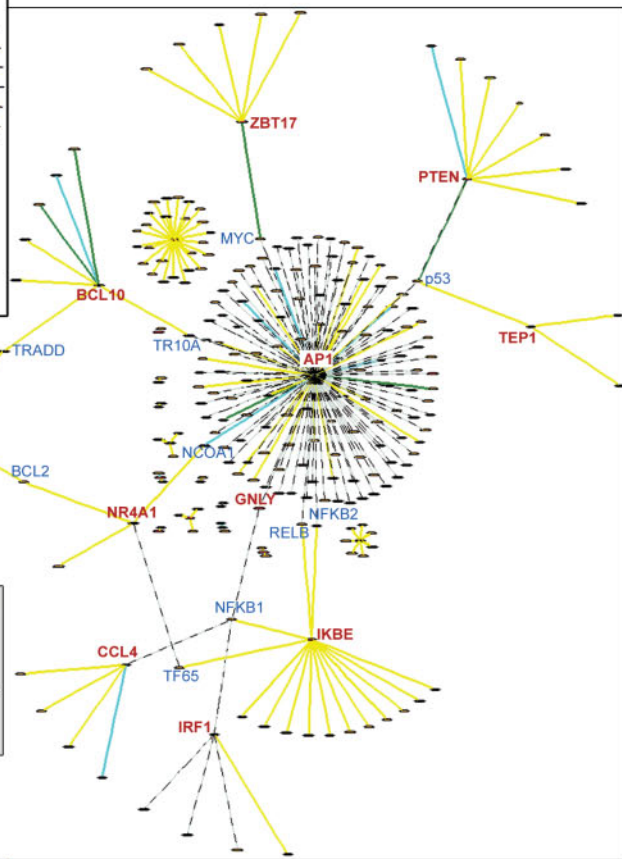
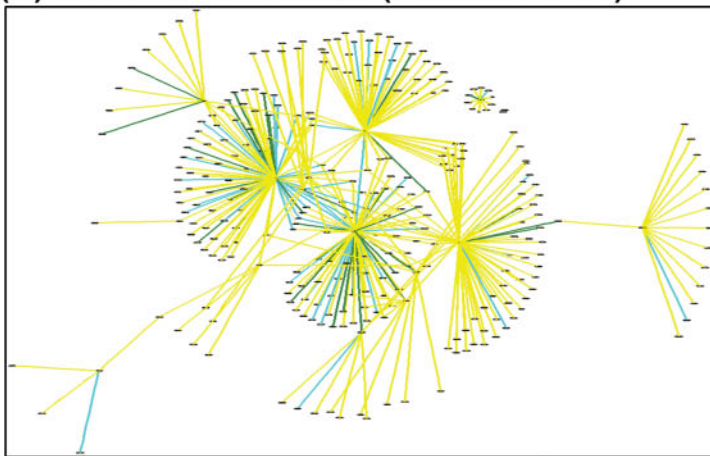
HUMAN-MODEL ORGANISMS SIMILARITY SEARCH (Reciprocal Best Matches using Blastp)

QUERY PROTEIN	UNIPROT	SPECIES	%IDENTITY	ALIGNMENT	SCORE	E-VALUE	INTERACTION
Transforming protein p21	P21_HUMAN	Homo sapiens					
Transforming protein p21	P21_MOUSE	Mus musculus	98.94	358	189	2.00E-99	
Transforming protein p21	P21_RAT	Rattus norvegicus	98.94	358	189	6.00E-100	Interaction info unavailable

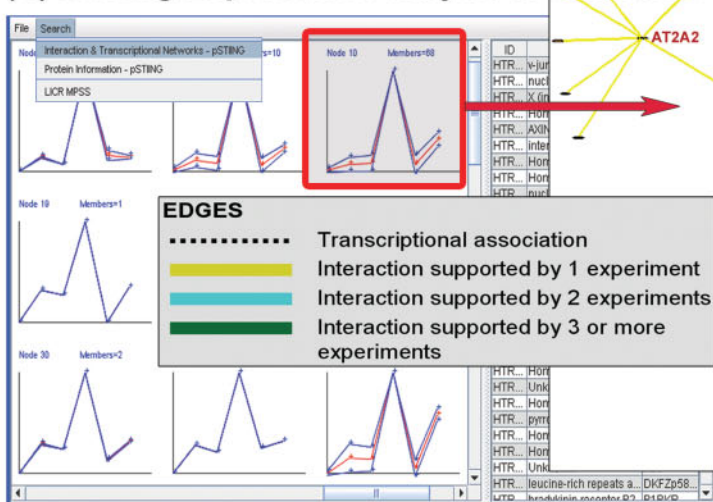
QUERY PROTEIN	UNIPROT	SPECIES	%IDENTITY	ALIGNMENT	SCORE	E-VALUE	INTERACTION
80 kDa heat shock protein, mitochondrial precursor	HSP60_HUMAN	Homo sapiens					
80 kDa heat shock protein, mitochondrial precursor	HSP60_DROME	Drosophila melanogaster	73.92	793	556	0	
80 kDa heat shock protein, mitochondrial precursor	HSP60_MOUSE	Mus musculus	97.48	1014	556	0	Interaction info unavailable
Heat shock protein 80, mitochondrial precursor	HSP60_YEAST	Saccharomyces cerevisiae	56.85	580	540	2.00E-168	Interaction info unavailable
Chaperonin homolog HSP60, mitochondrial precursor	HSP60_CAEEL	Caenorhabditis elegans	71.69	736	544	0	Interaction info unavailable
80 kDa heat shock protein, mitochondrial precursor	HSP60_RAT	Rattus norvegicus	97.48	1014	556	0	Interaction info unavailable



(B) Networks in Disease (Breast Cancer)



(C) Linking Expression Analysis to Networks



known for a protein onto homologous (and where possible, orthologous) proteins across species in which the interaction information is not as extensive or not available (Figure 2A). Our aim is to obtain as much additional insights as it is possible about the human interactome by inferring from model organisms.

pSTIING provides three approaches whereby users can choose to query interaction information across species. The first uses orthology groupings from Clusters of Orthologous Groups (COG/KOG) (19). The second approach involves UniRef 50/90 (2) clusters of proteins with at least 50 or 90% identity, respectively. The third employs a reciprocal best hit approach to map protein interaction information (where available) in model organisms (rat, mouse, fly, worm and yeast) onto human orthologues. Sequence similarity (BLASTP) searches against human proteins were performed for each model organism protein, and reciprocal searches carried out for each top BLAST hit. Each reciprocal best match is selected as a potential orthologue and its interactions displayed graphically. Also shown is the extent of sequence similarity (percent identity, BLAST score and *E*-value) between orthologous proteins, which provides an indication of the level of confidence associated with the inferred or projected interactions (Figure 2A and Supplementary Data 4).

INTERACTION NETWORKS VIA ONTOLOGY, MOTIFS AND INTERPRO DOMAINS

GO annotation (18) provides a well-structured set of controlled common vocabularies to describe the molecular function of gene products and proteins, the biological processes with which they are associated and the cellular compartments in which they are located. The inclusion of GO annotation terms and category identifiers into pSTIING facilitates the construction of interaction networks associated with a chosen GO annotation.

For example, a search using a phrase for a biological process like 'regulation of apoptosis', returns several possible matching terms, including 'positive regulation of apoptosis' (GO:0043065) and 'negative regulation of apoptosis' (GO:0043066) from which users could refine and select the appropriate term (Supplementary Data 5). There are several display options available. The 'protein' option generates graphical maps of interactions between proteins associated with the selected GO term. As for 'UniRef 50/90' and 'COG/KOG' options, the GO search is combined with cross-species homology information, producing graphical representations of interactions between UniRef or COG/KOG clusters, respectively.

It is also possible to explore interaction networks of proteins with similar InterPro domains or motifs. pSTIING

supports searches by InterPro accession, InterPro name or regular expression specifying conserved motifs (Supplementary Data 6).

HUMAN DISEASE AND INTERACTION NETWORKS

Many human diseases including inflammatory and autoimmune disorders (21–23), atherosclerosis (24) and cancers (25,26) have complex aetiologies associated with a variety of environmental cues feeding into a number of interaction networks or modules leading to aberrant perturbations at the 'systems' level.

A list of diseases was selected from the OMIM (20) morbid map initiative, and the information about proteins whose de-regulation or mutation have association or causal implication in disease was integrated with interaction information in pSTIING. For example, searching 'colorectal cancer' or 'breast cancer' will return information and OMIM links about the disorder (such as clinical synopses) and associated chromosomal loci. pSTIING will graphically represent interaction networks of proteins implicated in the disease (Figure 2B).

LINKING GENE EXPRESSION AND PROTEOMIC EXPERIMENTS TO INTERACTION AND REGULATORY NETWORKS

Filtered and normalized datasets from gene expression [including microarrays and massively parallel signature sequencing (MPSS)] and proteomic experiments can be imported into pSTIING via CLADIST, a clustering tool associated with pSTIING. pSTIING-CLADIST features various clustering algorithms including hierarchical clustering, *k*-means and self-organizing maps (SOM) from which users could select the clusters of co-regulated genes or proteins and directly cross-correlate these with molecular interaction and transcriptional information in pSTIING.

This feature allows the contextual projection of co-expression patterns onto prior network information, facilitating the assembly of protein interaction and transcriptional regulatory networks into functional modules. For example, Figure 2C shows the identification of a set of genes that are co-ordinately expressed across tumour and non-tumour cells. When linked to information in pSTIING, preliminary data reveals an underlying network of protein–protein interactions and transcriptional regulatory associations, suggesting their involvement in the same functional module or programme within the context of the physiological state examined.

Figure 2. (A) The identification of potential orthologues using human-model organism reciprocal similarity (BLASTP) searches allows additional interaction information to be projected from model organism proteins onto orthologous proteins in human. The sequence similarity between orthologous proteins (percent identity, BLAST score and *E*-value) for each query protein is displayed alongside their interaction information and when consolidated across species, these interactions are mapped onto human proteins. As a comparison, the corresponding interaction map without orthology mapping is shown. (B) pSTIING supports querying by human disease, facilitating the exploration of interaction and transcriptional regulatory associations between proteins implicated in disease (e.g. breast cancer). (C) Linking of expression analysis to molecular interaction networks (represented by solid edges) and transcriptional regulatory modules (dashed lines) using pSTIING-CLADIST. The CLADIST tool associated with pSTIING provides various clustering algorithms including SOM, which was used to identify a cluster of co-ordinately expressed genes (represented by nodes with red labels) from MPSS expression analyses of tumour and non-tumour samples. These were then cross-correlated directly with interaction and transcriptional information in pSTIING to detect any underlying physical or functional associations.

FUTURE DEVELOPMENTS

Building on molecular interaction data provided by publicly available databases [including MINT, MIPS, BIND, IntAct (27), STRING (28) and DIP (29)], pSTIING offers a unique resource in that it integrates transcriptional regulatory associations with molecular interaction data and signalling modules, allowing researchers and students to explore and extend regulatory networks using experimental data or by searching on protein domains/motifs, GO annotations, human diseases and orthologous proteins across species.

We will continue to assemble molecular interaction and transcriptional regulatory networks into functional pathways, particularly those that underpin inflammatory processes, cell migration and tumourigenesis. Hence, we welcome and encourage contributions from investigators towards this open resource via the pSTIING website.

We will also expand the knowledgebase to include quantitative data from directed expression and functional studies to examine the temporo-spatial dynamics of regulatory networks as determined by cell type, the state of a cell, its micro-environment and extracellular cues, in the context of a specific physiological process or pathological condition. This is critical for the development of realistic and predictive quantitative models of biological systems, which we envisage pSTIING to facilitate in the future.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors wish to thank Anne Ridley and Buzz Baum for useful discussions and critical reading of the manuscript. pSTIING is a project within the MAIN-NOE (Network of Excellence for Cell migration in Chronic Inflammation) consortium supported and funded by the EU 6th Framework Programme (EU project no. FP6-502935). Funding to pay the Open Access publication charges for this article was provided by the Ludwig Institute for Cancer Research.

Conflict of interest statement. None declared.

REFERENCES

- Wheeler,D.L., Barrett,T., Benson,D.A., Bryant,S.H., Canese,K., Church,D.M., DiCuccio,M., Edgar,R., Federhen,S., Helmberg,W. *et al.* (2005) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **33**, D39–D45.
- Bairoch,A., Apweiler,R., Wu,C.H., Barker,W.C., Boeckmann,B., Ferro,S., Gasteiger,E., Huang,H., Lopez,R., Magrane,M. *et al.* (2005) The Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **33**, D154–D159.
- Bateman,A., Coin,L., Durbin,R., Finn,R.D., Hollich,V., Griffiths-Jones,S., Khanna,A., Marshall,M., Moxon,S., Sonnhammer,E.L. *et al.* (2004) The Pfam protein families database. *Nucleic Acids Res.*, **32**, D138–D141.
- Mulder,N.J., Apweiler,R., Atwood,T.K., Bairoch,A., Bateman,A., Binns,D., Bradley,P., Bork,P., Bucher,P., Cerutti,L. *et al.* (2005) InterPro, progress and status in 2005. *Nucleic Acids Res.*, **33**, D201–D205.
- Pearl,F., Todd,A., Sillitoe,I., Dibley,M., Redfern,O., Lewis,T., Bennett,C., Marsden,R., Grant,A., Lee,D. *et al.* (2005) The CATH Domain Structure Database and related resources Gene3D and DHS provide comprehensive domain family information for genome analysis. *Nucleic Acids Res.*, **33**, D247–D251.
- Zanzoni,A., Montecchi-Palazzi,L., Quondam,M., Ausiello,G., Helmer-Citterich,M. and Cesareni,G. (2002) MINT: a Molecular Interaction database. *FEBS Lett.*, **513**, 135–140.
- Alfarano,C., Andrade,C.E., Anthony,K., Bahroos,N., Bajec,M., Bantoft,K., Betel,D., Bobecko,B., Boutilier,K., Burgess,E. *et al.* (2005) The Biomolecular Interaction Network Database and related tools 2005 update. *Nucleic Acids Res.*, **33**, D418–D424.
- Kanehisa,M., Goto,S., Kawashima,S., Okuno,Y. and Hattori,M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.*, **32**, D277–D280.
- Li,S., Armstrong,C.M., Bertin,N., Ge,H., Milstein,S., Boxem,M., Vidalain,P.O., Han,J.D., Chesneau,A., Hao,T. *et al.* (2004) A map of the interactome network of the metazoan *C.elegans*. *Science*, **303**, 540–543.
- Zhu,H., Bilgin,M., Bangham,R., Hall,D., Casamayor,A., Bertone,P., Lan,N., Jansen,R., Bidlingmaier,S., Houfek,T. *et al.* (2001) Global analysis of protein activities using proteome chips. *Science*, **293**, 2101–2105.
- Gavin,A.C., Bosche,M., Krause,R., Grandi,P., Marzioch,M., Bauer,A., Schultz,J., Rick,J.M., Michon,A.M., Cruciat,C.M. *et al.* (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, **415**, 141–147.
- Bouwmeester,T., Bauch,A., Ruffner,H., Angrand,P.O., Bergamini,G., Crougton,K., Cruciat,C., Eberhard,D., Gagneur,J., Ghidelli,S. *et al.* (2004) A physical and functional map of the human TNF- α /NF- κ B signal transduction pathway. *Nature Cell Biol.*, **6**, 97–105.
- Ho,Y., Gruhler,A., Heilbut,A., Bader,G.D., Moore,L., Adams,S.L., Millar,A., Taylor,P., Bennett,K., Boutilier,K. *et al.* (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature*, **415**, 180–183.
- Kitano,H. (2002) Computational systems biology. *Nature*, **420**, 206–210.
- Hartwell,L.H., Hopfield,J.J., Leibler,S. and Murray,A.W. (1999) From molecular to modular cell biology. *Nature*, **402**, C47–C52.
- Zhao,F., Xuan,Z., Liu,L. and Zhang,M.Q. (2005) TRED: a Transcriptional Regulatory Element Database and a platform for in silico gene regulation studies. *Nucleic Acids Res.*, **33**, D103–D107.
- Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
- Camon,E., Magrane,M., Barrell,D., Lee,V., Dimmer,E., Maslen,J., Binns,D., Harte,N., Lopez,R. and Apweiler,R. (2004) The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with gene ontology. *Nucleic Acids Res.*, **32**, D262–D266.
- Tatusov,R.L., Fedorova,N.D., Jackson,J.D., Jacobs,A.R., Kiryutin,B., Koonin,E.V., Krylov,D.M., Mazumder,R., Mekhedov,S.L., Nikolskaya,A.N. *et al.* (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*, **4**, 41.
- Hamosh,A., Scott,A.F., Amberger,J.S., Bocchini,C.A. and McKusick,V.A. (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **33**, D514–D517.
- Ollier,W.E. (2004) Cytokine genes and disease susceptibility. *Cytokine*, **28**, 174–178.
- Aderem,A. and Smith,K.D. (2004) A systems approach to dissecting immunity and inflammation. *Semin. Immunol.*, **16**, 55–67.
- Lehner,B., Semple,J.I., Brown,S.E., Counsell,D., Campbell,R.D. and Sanderson,C.M. (2004) Analysis of a high-throughput yeast two-hybrid system and its use to predict the function of intracellular proteins encoded within the human MHC class III region. *Genomics*, **83**, 153–167.
- King,J.Y., Ferrara,R., Tabibiazar,R., Spin,J.M., Chen,M.M., Kuchinsky,A., Vailaya,A., Kincaid,R., Tsalenko,A., Deng,D.X. *et al.* (2005) Pathway analysis of coronary atherosclerosis. *Physiol Genomics*, **23**, 103–118.
- Lin,B., White,J.T., Lu,W., Xie,T., Uteg,A.G., Yan,X., Yi,E.C., Shannon,P., Khrebtukova,I., Lange,P.H. *et al.* (2005) Evidence for the presence of disease-perturbed networks in prostate cancer cells by genomic and proteomic analyses: a systems approach to disease. *Cancer Res.*, **65**, 3081–3091.
- Khalil,I.G. and Hill,C. (2005) Systems biology for cancer. *Curr. Opin. Oncol.*, **17**, 44–48.
- Hermjakob,H., Montecchi-Palazzi,L., Lewington,C., Mudali,S., Kerrien,S., Orchard,S., Vingron,M., Roechert,B., Roepstorff,P.,

- Valencia, A. *et al.* (2004) IntAct: an open source molecular interaction database. *Nucleic Acids Res.*, **32**, D452–D455.
28. von Mering, C., Jensen, L.J., Snel, B., Hooper, S.D., Krupp, M., Foglierini, M., Jouffre, N., Huynen, M.A. and Bork, P. (2005) STRING: known and predicted protein–protein associations, integrated and transferred across organisms. *Nucleic Acids Res.*, **33**, D433–D437.
29. Salwinski, L., Miller, C.S., Smith, A.J., Pettit, F.K., Bowie, J.U. and Eisenberg, D. (2004) The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.*, **32**, D449–D451.
30. Maglott, D., Ostell, J., Pruitt, K.D. and Tatusova, T. (2005) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **33**, D54–D58.
31. Wain, H.M., Lush, M.J., Ducluzeau, F., Khodiyar, V.K. and Povey, S. (2004) Genew: the Human Gene Nomenclature Database, 2004 updates. *Nucleic Acids Res.*, **32**, D255–D257.
32. Joshi-Tope, G., Gillespie, M., Vastrik, I., D’Eustachio, P., Schmidt, E., de Bono, B., Jassal, B., Gopinath, G.R., Wu, G.R., Matthews, L. *et al.* (2005) Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.*, **33**, D428–D432.
33. Peri, S., Navarro, J.D., Kristiansen, T.Z., Amanchy, R., Surendranath, V., Muthusamy, B., Gandhi, T.K., Chandrika, K.N., Deshpande, N., Suresh, S. *et al.* (2004) Human protein reference database as a discovery resource for proteomics. *Nucleic Acids Res.*, **32**, D497–D501.
34. Fitzgerald, K.A., O’Neill, L.A.J., Gearing, A.J.H. and Callard, R.E. (2001) *The Cytokine Factsbook*, 2nd edn. Academic Press, San Diego.
35. Isacke, C.M. and Horton, M.A. (2000) *The Adhesion Molecule Factsbook*, 2nd edn. Academic Press, San Diego.
36. Pagel, P., Kovac, S., Oesterheld, M., Brauner, B., Dunger-Kaltenbach, I., Frishman, G., Montrone, C., Mark, P., Stumpflen, V., Mewes, H.W. *et al.* (2005) The MIPS mammalian protein–protein interaction database. *Bioinformatics*, **21**, 832–834.