

Wiki-Pi: A Web-Server of Annotated Human Protein-Protein Interactions to Aid in Discovery of Protein Function

Naoki Orii^{1,2}, Madhavi K. Ganapathiraju^{1*}

1 Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America, **2** Language Technologies Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States of America

Abstract

Protein-protein interactions (PPIs) are the basis of biological functions. Knowledge of the interactions of a protein can help understand its molecular function and its association with different biological processes and pathways. Several publicly available databases provide comprehensive information about individual proteins, such as their sequence, structure, and function. There also exist databases that are built exclusively to provide PPIs by curating them from published literature. The information provided in these web resources is protein-centric, and not PPI-centric. The PPIs are typically provided as lists of interactions of a given gene with links to interacting partners; they do not present a comprehensive view of the nature of both the proteins involved in the interactions. A web database that allows search and retrieval based on biomedical characteristics of PPIs is lacking, and is needed. We present Wiki-Pi (read Wiki- π), a web-based interface to a database of human PPIs, which allows users to retrieve interactions by their biomedical attributes such as their association to diseases, pathways, drugs and biological functions. Each retrieved PPI is shown with annotations of both of the participant proteins side-by-side, creating a basis to hypothesize the biological function facilitated by the interaction. Conceptually, it is a search engine for PPIs analogous to PubMed for scientific literature. Its usefulness in generating novel scientific hypotheses is demonstrated through the study of IGSF21, a little-known gene that was recently identified to be associated with diabetic retinopathy. Using Wiki-Pi, we infer that its association to diabetic retinopathy may be mediated through its interactions with the genes HSPB1, KRAS, TMSB4X and DGKD, and that it may be involved in cellular response to external stimuli, cytoskeletal organization and regulation of molecular activity. The website also provides a wiki-like capability allowing users to describe or discuss an interaction. Wiki-Pi is available publicly and freely at <http://severus.dbmi.pitt.edu/wiki-pi/>.

Citation: Orii N, Ganapathiraju MK (2012) Wiki-Pi: A Web-Server of Annotated Human Protein-Protein Interactions to Aid in Discovery of Protein Function. PLoS ONE 7(11): e49029. doi:10.1371/journal.pone.0049029

Editor: Peter Csermely, Semmelweis University, Hungary

Received: June 16, 2012; **Accepted:** October 3, 2012; **Published:** November 28, 2012

Copyright: © 2012 Orii and Ganapathiraju. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the Biobehavioral Research Awards for Innovative New Scientists (BRAINS) grant from the National Institute of Mental Health of the National Institutes of Health, United States of America (grant number MH094564 to MG). The grant information may be viewed at http://projectreporter.nih.gov/project_info_results.cfm?aid=8304920&icde=12774120. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: madhavi@pitt.edu

Introduction

Annotations of proteins such as their sequence, structure, interactions and functions, or their association to diseases and drugs, are provided by a number of web-based databases such as Uniprot [1], HPRD [2], Gene Cards [3], Gene Ontology [4], KEGG [5], PDB [6], OMIM [7] and REACTOME [8]. Some databases such as BioGRID [9], STRING [10], DIP [11], MINT [12], InnateDB [13], and IntAct [14] are designed exclusively to provide information about protein-protein interactions (PPIs). These PPI databases provide a valuable resource by curating experimentally known interactions, and have become the gold-standard data sources for a number of bioinformatic studies such as prediction of protein-protein interactions and protein functions, gene prioritizations and other systems biology studies. The contribution of most of these websites is the presentation of datasets that are painstakingly compiled by curators from literature. Conversely, a *crowdsourcing* model for curating protein annotations was explored by WikiGenes [15]. Similar to

Wikipedia, users can collaboratively create, edit and update articles on the site. Thus, instead of a small group of creators, researchers around the globe are able to contribute to that knowledge base. However, all of these web-based data resources provide a gene-centric view of interactions. That is, the “central players” of these databases are genes and not the interactions. In most of these web resources, interactions are merely provided as lists with respect to a specific protein, and any information about the interactions, if provided, is about the type of interaction or the experimental method or publication that reports the said interaction. Although the information that an interaction exists between two proteins is useful by itself, it may be insufficient from a biomedical researcher’s perspective. Biomedical researchers often have one or a few proteins that they study in detail, and exploring the interactions of these proteins requires rich annotations about the interacting partners in order to identify an interaction that is relevant to their research – namely, an interaction that would potentially lead to further experiments in their own lab.

ATF1 and BRCA1

Number of citations of the paper that reports this interaction (PubMedID [10945975](#)) as of Mar 2012:

6

Data Source:

[BioGRID](#) (affinity chromatography technology)

[HPRD](#) (in vitro, in vivo, two hybrid)

Discuss ↓

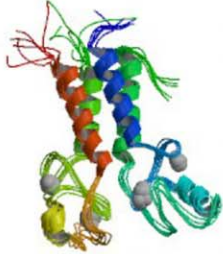
	ATF1	BRCA1
Description	activating transcription factor 1	breast cancer 1, early onset
Image	No pdb structure	
Cellular Component	<ul style="list-style-type: none"> Nucleus Nucleoplasm Transcription Factor Complex 	<ul style="list-style-type: none"> Ubiquitin Ligase Complex Condensed Nuclear Chromosome Ruffle Intracellular Nucleus Nucleoplasm Cytoplasm Mitochondrial Matrix Plasma Membrane Focal Adhesion Gamma-tubulin Ring Complex Ribonucleoprotein Complex BRCA1-BARD1 Complex Filamentous Actin Protein Complex BRCA1-A Complex
Molecular Function	<ul style="list-style-type: none"> Sequence-specific DNA Binding Transcription Factor Activity Sequence-specific Distal Enhancer Binding RNA Polymerase II Transcription Factor Activity Protein Complex Binding Sequence-specific DNA Binding 	<ul style="list-style-type: none"> DNA Binding Damaged DNA Binding Transcription Coactivator Activity RNA Binding Ubiquitin-protein Ligase Activity Protein Binding Zinc Ion Binding Tubulin Binding Ligase Activity

Figure 1. PPI page. A PPI page showing annotations about the proteins and about the interaction. URL: <http://severus.dbmi.pitt.edu/wiki-pi/index.php/pair/view/466/672>. doi:10.1371/journal.pone.0049029.g001

Currently there is no search engine that allows retrieval of PPIs by their biomedical associations. Existing databases primarily allow a user to search for interactions by gene symbol or other widely used identifiers, be it protein/gene name, Entrez gene identifier, or Ensembl identifier. However, biologists specializing in the study of a certain disease or pathway may be interested in retrieving interactions associated with that disease or pathway, and not by a single gene. For example, a researcher studying diabetes is not able to retrieve PPIs associated with diabetes using any of the existing PPI databases (although specialized databases may exist occasionally for a few well-studied diseases). InnateDB and IntAct provide search functionality, and users can search for PPIs by experimental details but not by specifying biomedical attributes of the proteins.

PPIs can contribute to the discovery of a gene's biological function. An example where PPIs have contributed to the discovery of gene function is *Disrupted in Schizophrenia 1* (DISC1), a novel protein discovered in 2000 with no known homolog in human. DISC1 was identified to be associated with schizophrenia; although it had well characterized protein domains such as coiled-coil domains, leucine-zipper domains, and nuclear localization and export signals, nothing was inferred about its function [16,17]. To understand the function of DISC1, PPIs were determined using yeast 2-hybrid technology [18,19]. Availability of this 'DISC1 interactome' has led to a large number of studies that concluded the association of DISC1 to cAMP signaling, axon elongation and neuronal migration, and accelerated the research pertaining to schizophrenia in general and DISC1 in particular [20]. Therefore,

Wiki-Pi Home About Search Advanced Search

Wiki-Pi

Search for protein-protein interactions:

e.g. EGFR, cancer, apoptosis Search Advanced Search

Interactions related with HIV

CCNT1 and HTATSF1
CDK9 and HTATSF1
ETS1 and HTATIP2
GTF2F2 and HTATSF1
NEUROG1 and TARBP2
NF2 and VPRBP
NF2 and TARBP2
POLR2A and HTATSF1
EIF2AK2 and TARBP2
RAG1 and VPRBP

View more »

Frequently Viewed Interactions

SERPINA3 and ELANE
ATF1 and BRCA1
ACTA1 and EGFR
CCDC85B and AGGF1
SOX9 and CREB3L4
ID3 and IFI16
ITGB1 and NF2
MDM2 and RPS27A
PDGFRB and STAT5A
POLR2C and UBE2W

Random Interactions

AQP4 and CSNK2A1
CDK4 and NCOA2
F11 and HGF
FLT3 and FLT3LG
GNAQ and KDR
GTF2B and GTF2H4
SMAD4 and NEK6
MAGEA11 and C18orf54
RAB27A and SYTL1
SNAPC1 and SNAPC5

© Madhavi K. Ganapathiraju 2012

Figure 2. Website homepage. The homepage gives a search box, and also shows a shortlist of interactions some of which are populated randomly from the database while others are those that are most-frequently searched on Wiki-Pi. URL: <http://severus.dbmi.pitt.edu/wiki-pi/>. doi:10.1371/journal.pone.0049029.g002

it is useful to have a web resource of PPIs that displays not only the symbols of interacting partners but also comprehensive information on what the interacting partners of a gene can tell about the gene itself.

We developed a web resource, Wiki-Pi, which addresses the above issues. It provides an effective means to search and retrieve interactions of interest, and displays the retrieved interactions with annotations of their biomedical associations so as to enable further discoveries. The search for interactions can be carried out by

Table 1. Data sources.

Data type	Data source	Link
Interactions	HPRD	http://www.hprd.org/ (BINARY_PROTEIN_PROTEIN_INTERACTIONS.txt)
Interactions	BioGRID	http://thebiogrid.org/downloads/archives/Latest%20Release/BIOGRID-ALL-LATEST.mitab.zip
Entrez ID	NCBI Gene	ftp://ftp.ncbi.nih.gov/gene/DATA/gene_info.gz
GO annotations	NCBI Gene2go	ftp://ftp.ncbi.nih.gov/gene/DATA/gene2go.gz
Ensembl IDs	NCBI Gene2ensembl	ftp://ftp.ncbi.nih.gov/gene/DATA/gene2ensembl.gz
Uniprot IDs	UniProt	ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/idmapping/by_organism/HUMAN_9606_idmapping.dat.gz
List of drugs binding to a protein	DrugBank	http://www.drugbank.ca/system/downloads/current/drugbank.txt.zip
List of pathways	REACTOME	http://www.reactome.org/download/current/uniprot_2_pathways.stid.txt
List of diseases	KEGG	http://www.genome.jp/kegg/
Pubmed articles citing a gene and their abstracts	pubmed2ensembl	http://www.pubmed2ensembl.org/
GO Term enrichment computation	BiNGO	http://www.psb.ugent.be/cbd/papers/BiNGO/Customize.html

doi:10.1371/journal.pone.0049029.t001

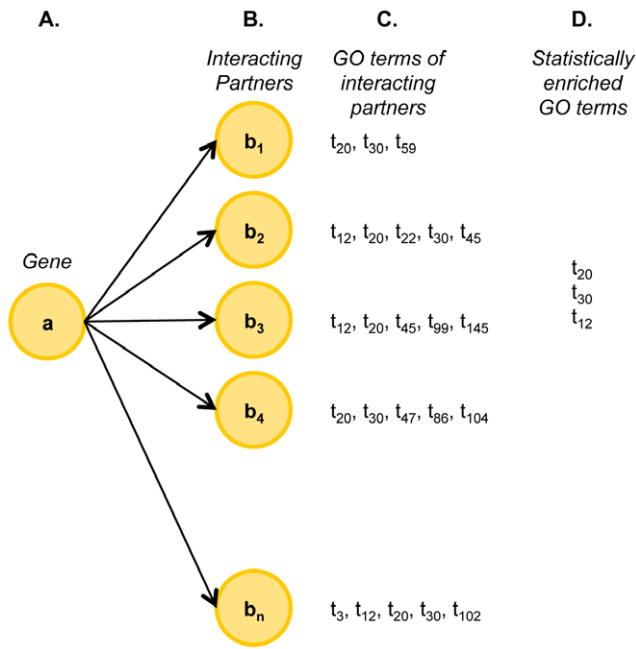


Figure 3. Concept diagram of GO term enrichment calculation. Gene *a* interacts with genes b_1, \dots, b_n . GO terms t_i of each interacting partner are shown to its right. BiNGO computes the statistically enriched GO terms (functional categories that the genes are enriched in, and find that the statistically enriched GO terms are t_{20}, t_{30} , and t_{12} . See methods in [24] for details of computation. doi:10.1371/journal.pone.0049029.g003

specifying biological and disease-relevant annotations of genes. Wiki-Pi provides the seed information necessary for gene function discoveries, by readily presenting the annotations of the gene at hand as well as those of its interacting partners. Further, Wiki-Pi facilitates *knowledge-creation via crowdsourcing*. It allows users to discuss or describe their hypothesis, or other known facts that are not part of existing database, in the wiki portion of each interaction. The website is freely available at <http://severus.dbmi.pitt.edu/wiki-pi> and is viewable in all major browsers including those on smartphones and e-readers.

Data and Functionality

Wiki-Pi is a web resource whose focus is on telling the story of each interaction in the human interactome. Only binary biophysical interactions are presented. Each interaction can be viewed on its own webpage (Figure 1). The mechanism to reach individualized PPI pages is via the search functionality provided on the homepage (Figure 2) or via a search box provided conveniently at the top of any page.

Data Sources

Binary biophysical interactions of the human interactome have been collected from HPRD and BioGRID. Currently, Wiki-Pi contains 48,419 unique interactions among 10,492 proteins. Data sources for annotations are given in Table 1. Excluding HPRD, all of the data from the databases is automatically updated monthly. Only data from HPRD is updated manually (we note that HPRD has not updated its database since April 13, 2010). We rely on these databases for curated PPIs, and do not curate them from other resources ourselves. The database of interactions and other annotations are loaded into MySQL.

Individualized Page for Each PPI

A webpage of a PPI consists of two sections: an *automatically generated annotation section* with detailed annotations describing the interaction and its participant proteins, and a *wiki section* where users can discuss the interaction. The details of the annotation section from top to bottom are as follows (see Figure 1).

Biomedical Annotations. The top of the section gives a link to the PubMed record of the original publication reporting the interaction; this publication source is obtained from HPRD or BioGRID. Following that, the count of papers citing that publication is shown; this count is obtained from PubMed. The citation count is provided so as to give an idea of the extent of the scientific impact of that interaction. Sometimes the original publication is cited more for the experimental method than for the interactions itself, but this can be easily concluded by following the PubMed link to the original publication. Next, biologically and medically relevant characteristics of the two participant proteins are shown where available: PDB IDs and structure, Gene Ontology cellular component, molecular function and biological process terms at the GO Slim level, associated pathways from REACTOME, associated diseases from KEGG, and drugs binding to that protein from DrugBank [21]. These annotations provide useful information for analyzing the biological function of the given interaction. Additionally, links to corresponding pages of the genes in other databases, namely, Entrez gene [22], HPRD, Ensembl [23], and Uniprot, are provided.

GO Terms Enriched among Interacting Partners. A unique feature of this web resource is that it provides for each gene in the interaction, a list of Gene Ontology biological process terms *statistically enriched* among its interacting partners. The enriched terms are computed by employing BiNGO plugin in Cytoscape [24,25]. The hypergeometric statistical test of significance is used with a Benjamini & Hochberg False Discovery Rate (FDR) correction at a significance level of 0.05. For instance, when calculating enriched terms for gene ‘a’ (see Figure 3), the study group consists of the interacting partners b_1, b_2, \dots, b_n , while the reference set consists of n genes randomly selected from the entire genome. BiNGO then collects GO biological process terms of b_1, b_2, \dots, b_n . For each of the terms in the collection, it computes whether the number of genes associated with that term is significantly greater among interacting partners compared to that of random set. The methodology is described in detail in the original publication of BiNGO [24]. For a given gene (‘a’), if more than 50 terms are found to be enriched among interacting partners associations, only the top 50 enriched terms in the order of increasing p-value or decreasing statistical significance are shown on the website. For example, when viewing the annotations for an interaction between DISC1 and another protein, GO biological process terms that are significantly overrepresented in DISC1’s interacting partners are shown. Viewing these terms would provide a handle for biologists in determining any novel associations of that gene in specific biological processes or diseases. These terms are especially useful when many interactions are known for a protein, but its functional characteristics are unknown [26].

Tag Clouds from Abstracts. To give an overview of the topics that each of these genes are associated with, *tag clouds* are presented which are constructed from abstracts of papers associated with each protein as given by pubmed2ensembl [26]. An interaction may be more interesting if it connects two different processes together, whereas it may be less novel if the interaction is between two proteins which participate in same biological process. Therefore, in addition to the above tag clouds, another tag cloud is displayed for each protein made up of words that associate with

The image shows two screenshots of the Wiki-Pi website. The top screenshot displays the 'Advanced Search' form with fields for 'Query gene' and 'Interacting partner (optional)'. The bottom screenshot shows the 'Advanced Search Results' page for the query 'all these words: e.g. EGFR, apoptosis' and 'Interacting partner (optional): e.g. disease:"lung cancer"'. The results table shows 517 interactions found, sorted by Relevance. The first result is for the interaction between BRCA1 and STAT3. The table columns are: Symbols, Name 1, Name 2, Pathways 1, Pathways 2, Drugs 1, Drugs 2, Diseases 1, and Diseases 2. The 'Diseases 2' column for this interaction lists: Breast cancer, Ovarian cancer, and a list of other well-defined immunodeficiency syndromes including Wiskott-Aldrich syndrome, DiGeorge syndrome, Hyper-IgE syndrome, X-linked lymphoproliferative syndrome, Immunodeficiency, Polyendocrinopathy, Enteropathy, X-linked Syndrome (IPEX), Cartilage-Hair Hypoplasia, Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), and Oral cancer.

Advanced Search
Find interactions with...

Query gene: e.g. EGFR, apoptosis
Interacting partner (optional): e.g. disease:"lung cancer"

all these words: e.g. EGFR, apoptosis
any of these words: e.g. AKT1 DISC1 BDNF
none of these words:

Search

Advanced Search Results
517 interactions found: Sort by: Relevance Number of Pathways Number of Drugs Number of Diseases

Symbols	Name 1	Name 2	Pathways 1 Hide	Pathways 2 Hide	Drugs 1 Hide	Drugs 2 Hide	Diseases 1 Hide	Diseases 2 Hide
BRCA1 and STAT3	breast cancer 1, early onset	signal transducer and activator of transcription 3 (acute-phase response factor)	<ul style="list-style-type: none"> Cell Cycle DNA Repair Meiosis 	<ul style="list-style-type: none"> Immune System Signal Transduction 			<ul style="list-style-type: none"> Breast cancer Ovarian cancer 	<ul style="list-style-type: none"> Other well-defined immunodeficiency syndromes, including the following seven diseases: Wiskott-Aldrich syndrome; DiGeorge syndrome; Hyper-IgE syndrome; X-linked lymphoproliferative syndrome; Immunodeficiency, Polyendocrinopathy, Enteropathy, X-linked Syndrome (IPEX); Cartilage-Hair Hypoplasia; Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) Oral cancer

Figure 4. Advanced-search feature. Image shows the results of the search where one gene is involved in the immunity pathway, while the other gene contains the term cancer anywhere in its annotation but not the word immunity. Note that the results can be sorted by number of pathways, diseases or drugs associated with the genes (counts of each gene are considered individually). URL: <http://severus.dbmi.pitt.edu/wiki-pi/index.php/search/adv?a-all=pathway%3Aimmunity&b-any=cancer&b-none=immunity>. doi:10.1371/journal.pone.0049029.g004

one protein but not the other. The tag cloud for a given gene is calculated as follows. First, a given gene's Ensembl identifier is mapped into a PMID (PubMed identifier) as given in `pubmed2ensembl` (<http://www.pubmed2ensembl.org/>) data. The abstract of the said publication is obtained, and it is treated as a document representing that particular gene. Starting with all of the abstracts as a corpus, stop words (such as 'for', 'it', 'the', etc) are removed, and stemming is carried out on the remaining words. Tf-idf, which is a measure of relevance used in information retrieval, is computed. *tf* refers to term-frequency and *idf* refers to inverse document-frequency, and *tf-idf* gives the relevance of a term to a given document ([27]). The size of a word in the tag cloud corresponds to the values of *tf-idf* for that term with respect to the document.

Wiki for Further Annotations by Users. The second section of the interaction page is the wiki, where users are encouraged to provide insights and discuss predictions about the

relevance of the interaction in a biological process, disease or pathway. The wiki section may be used for crowdsourcing not only knowledge *curation* but also knowledge *creation* about each interaction.

Navigation through Search

Users navigate Wiki-Pi primarily by using the search interface. Wiki-Pi allows full-text search as well as field-specific search; it does not require users to have the knowledge of any form of query language like Structured Query Language (SQL).

Indexing for Information Retrieval. The index for free-text search is constructed from gene symbols, gene names, GO annotations, pathways, drugs, and diseases (but not enriched GO terms and abstracts). Stop words are removed and stemming is carried out on all the content prior to indexing. Stemming in the context of information retrieval is a process by which words like

Search Results for: brca1

Can't find the interaction you're looking for? Try the [advanced search](#)

161 interactions involving **BRCA1** found:

Sort by:

Symbol

Number of Pathways

Number of Drugs

Number of Diseases

Interactant Symbol	Name	Associated Pathways Hide	Binding Drugs Hide	Associated Diseases Hide
BRCA1	breast cancer 1, early onset			
ABL1	c-abl oncogene 1, non-receptor tyrosine kinase	<ul style="list-style-type: none"> Developmental Biology Hemostasis 	<ul style="list-style-type: none"> Adenosine triphosphate Imatinib Dasatinib N-[4-Methyl-3-[[4-(3-Pyridinyl)-2-Pyrimidinyl]Amino]Phenyl]-3-Pyridinecarboxamide Nilotinib 2-[[[6-Oxo-1,6-dihydropyridin-3-yl]methylamino]-N-[4-propyl-3-(trifluoromethyl)phenyl]benzamide 1-[4-(pyridin-4-yloxy)phenyl]-3-[3-(trifluoromethyl)phenyl]urea MYRISTIC ACID 6-(2,6-dichlorophenyl)-2-[[3-(hydroxymethyl)phenyl]amino]-8-methylpyrido[2,3-d]pyrimidin-7(8h)-one 5-[3-(2-methoxyphenyl)-1H-pyrrolo[2,3-b]pyridin-5-yl]-N,N-dimethylpyridine-3-carboxamide 2-amino-5-[3-(1-ethyl-1H-pyrazol-5-yl)-1H-pyrrolo[2,3-b]pyridin-5-yl]-N,N-dimethylbenzamide 	<ul style="list-style-type: none"> Chronic myeloid leukemia (CML) Acute lymphoblastic leukemia (ALL) (precursor B lymphoblastic leukemia)
ACACA	acetyl-CoA carboxylase alpha	<ul style="list-style-type: none"> Metabolism 	<ul style="list-style-type: none"> Biotin 	
		<ul style="list-style-type: none"> Apoptosis Developmental Biology Disease Gene 	<ul style="list-style-type: none"> Adenosine triphosphate Arsenic trioxide 	<ul style="list-style-type: none"> PTEN hamartoma tumor syndrome (PHTS), including:

Figure 5. PPIs retrieved when searched by gene symbol. In these search results also, similar to those in Figure 3, the results can be sorted by number of pathways, diseases or drugs associated with the genes (counts of each gene are considered individually). URL: <http://severus.dbmi.pitt.edu/wiki-pi/index.php/search?q=brca1>. doi:10.1371/journal.pone.0049029.g005

‘inflammation’ and ‘inflammatory’ are mapped to their stem ‘inflamm’. When a word is queried, all interactions whose annotations (for either gene) contain that word are retrieved. The search functionality is created using the open-source search engine *Sphinx* (<http://sphinxsearch.com/>).

Search Functionality. The interactions may be retrieved with a simple search where any of the indexed content is given in the search box. For example, a query can be simply the gene symbol (e.g., AKT1) or any term that appears among the annotations of the gene (e.g. ‘blood’, ‘cytokine’, ‘hemostasis’). As stemming has been performed on all the words prior to indexing, searching for “inflammation” will retrieve interactions that have not only the word *inflammation* but also the word *inflammatory*. By allowing users to search for interactions based on fields such as GO terms, pathways, diseases, and drugs, researchers without a particular protein in mind can still successfully retrieve interactions of their interest. When multiple words are given in a simple search box, interactions containing all of the words are retrieved. An advanced-search page is also provided to retrieve interactions with more complex queries. Here, users can construct queries such as “DISC1 but not immunity”, “interactions of any of these proteins: TLR1, TLR2, ...”, “genes associated with schizophrenia that interact with genes associated with immunity” and so on. An example is shown in Figure 4, where the query is: “an interaction where one gene is involved in the immunity pathway, while the

other gene contains the term cancer anywhere in its annotation but not the word immunity”. Note that the users do not type such natural language sentences, but will type out query words in appropriate boxes in the advanced search page. Advanced search also allows users to restrict search to any of these fields: disease, pathway, drug, symbol, gene name, GO terms, or Entrez identifier (e.g. ‘disease:diabetes’, ‘pathway:hemostasis’ or ‘drug: diflunisal’).

Display of Search Results. The results of the search are presented in a tabular format showing gene symbols, names, pathways, diseases and drugs of the participant genes (Figures 5 and 6). The rows are sortable by the number of attributes associated with the genes. Each interaction may be clicked to view the detailed annotations page of the interaction (Figure 1).

Results and Discussion

Formulation of Novel Hypotheses Uniquely Enabled by Wiki-Pi

Unique features available in Wiki-Pi enable addressing scientific queries that are otherwise not feasible by other tools. Without Wiki-Pi, a biomedical scientist is left with manual curation of information from several data sources without a guarantee on finding the seed evidence required to crystallize a novel hypothesis. A comparison of functionality in Wiki-Pi and those of other existing PPI databases is given in Table 2. Note that Wiki-Pi is the

Table 2. Comparison of functionality of Wiki-Pi with other PPI databases.

	SEARCHING FOR INTERACTIONS				ANNOTATIONS ABOUT INTERACTIONS											
	Query on interaction characteristics	Query by characteristics of a gene	Query with constraints on characteristics of interacting partner	Results of a search	Details of each interaction	Side-by-side annotations of both proteins	Experimental Evidence	Publication reporting the interaction	Impact achieved by that interaction (number of citations of original publication)	GO annotations of both proteins	Pathways of both proteins	Drugs binding to each of the two proteins	Diseases associated with both proteins	Tag-cloud from abstracts for both proteins	Cross references to external databases	Crowdsourcing further discussion about each PPI
Wiki-Pi	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
STRING	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
IntACT	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
BioGRID	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
HPRD	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

The search functionality and annotations displayed for the retrieved interactions are compared across different PPI databases. For each function, the cell shows a tick mark if the function is supported by the corresponding webserver.
doi:10.1371/journal.pone.0049029.t002

Search Results for: alzheimers

Can't find the interaction you're looking for? Try the [advanced search](#)

189 interactions found:

Sort by:

Symbols	Name 1	Name 2	Pathways 1 Hide	Pathways 2 Hide	Drugs 1 Hide	Drugs 2 Hide	Diseases 1 Hide	Diseases 2
APP and PSEN2	amyloid beta (A4) precursor protein	presenilin 2 (Alzheimer disease 4)	<ul style="list-style-type: none"> Disease Hemostasis Immune System Signal Transduction 	<ul style="list-style-type: none"> A third proteolytic cleavage releases NICD Signal Transduction 			<ul style="list-style-type: none"> Alzheimer's disease (AD) 	<ul style="list-style-type: none"> Alzheimer disease (AD)
APOE and APP	apolipoprotein E	amyloid beta (A4) precursor protein	<ul style="list-style-type: none"> Metabolism 	<ul style="list-style-type: none"> Disease Hemostasis Immune System Signal Transduction 	<ul style="list-style-type: none"> Human Serum Albumin Serum albumin iodinated 		<ul style="list-style-type: none"> Lipoprotein glomerulopathy Hyperlipoproteinemia, type III; Dysbetalipoproteinemia Alzheimer's disease (AD) 	<ul style="list-style-type: none"> Alzheimer disease (AD)
								<ul style="list-style-type: none"> Acne invulgaris Hidradenoma suppurative Frontotemporal lobar degeneration (FTLD), including disease of brain; Frontotemporal dementia

Figure 6. PPIs retrieved when searched by disease. In these search results also, similar to those in Figure 3, the results can be sorted by number of pathways, diseases or drugs associated with the genes (counts of each gene are considered individually). URL: <http://severus.dbmi.pitt.edu/wiki-pi/index.php/search?q=alzheimers>. doi:10.1371/journal.pone.0049029.g006

sole database that allows a user to search by specifying conditions about *both* the proteins involved in a given interaction. Imposing strict conditions on the interaction in effect narrows down the search space of PPIs; this is critical, as there are tens of thousands of PPIs available in existing databases. This capability is invaluable when hypothesizing functions of genes that are not well-studied.

Wiki-Pi is especially useful today, as several genome-wide association studies (GWAS) are being published. GWAS studies are unbiased by current scientific knowledge (i.e. they do not have literature-bias) and often implicate genes with currently unknown biological functions to be associated with the disease under study. The number of GWAS studies has increased rapidly in the past couple of years. So far, 1,309 publications have reported GWAS results on 674 traits or diseases (www.genome.gov/gwastudies [28], accessed 2012-July-17). Though extensive work is being carried out to identify the common genetic variants that influence various diseases or traits through GWAS, the role of these genes and the exact mechanism of their action are yet to be discovered. Very little information is available about some of the GWAS-identified genes in terms of their molecular function and biological process. Wiki-Pi enables researching each of these genes and provides novel insights that may not otherwise materialize except when a scientist knows all the multiple specialized domains involved.

Possible Function of IGSF21 and the Likely Mediators of Its Association to Diabetic Retinopathy.

Using Wiki-Pi, we analyzed immunoglobulin superfamily member 21 (IGSF21) which has been identified through a recent GWAS study to be associated with diabetic retinopathy, where new blood vessels form at the back of the eye causing bleeding and blurring of vision [29]. There is no information currently known about IGSF21 except for the protein-protein interactions determined through high-throughput experiments and that it is an extracellular protein. Searching on Wiki-Pi for interactions of IGSF21, and then viewing the list of GO terms enriched among its interacting partners reveals that this extracellular protein may be involved in regulating metabolic processes, catalytic activity as well as cytoskeletal organization and response to external stimuli (see Figure 7 and File S1, generated by pasting list of interacting partners of IGSF21 into Cytoscape BiNGO plugin [24]). Although this enriched term calculation reveals that IGSF21 may be involved in signaling mechanisms in response to external stimuli, specifically in cytoskeletal organization, it does not reveal its relation to diabetic retinopathy. Its relation specifically to diabetic retinopathy is revealed further with the advanced-search feature of Wiki-Pi, which may be used to find interactions where one gene is IGSF21 and the other gene includes the term "blood" in any of its annotations (<http://severus.dbmi.pitt.edu/wiki-pi/index.php/search/adv?a-all=symbol%3Aigsf21&b-all=blood>). This query results in four interactions, namely

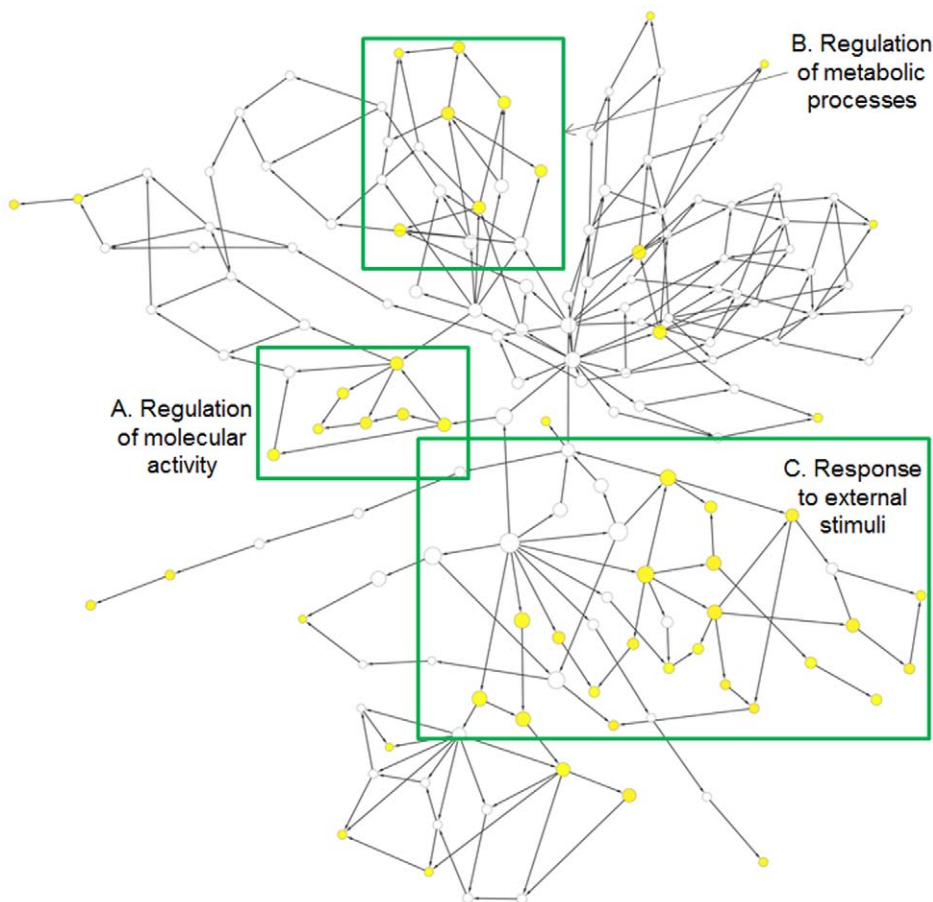


Figure 7. Statistically enriched Gene Ontology biological process terms of PPIs of IGSF21. Wiki-Pi website makes available only a list and not an image of enriched GO biological process terms. For clarification, this network diagram has been generated with BiNGO Cytoscape Plugin [24], for GO biological process terms, with the hypergeometric statistical test of significance, and a Benjamini & Hochberg False Discovery Rate (FDR) correction at a significance level of 0.05, by pasting the list of interacting partners (gene symbols) from Wiki-Pi. Statistical significance of the node (GO term) is shown in color, with the darker color indicating stronger significance. High-resolution image with labels of the nodes is available as File S1. doi:10.1371/journal.pone.0049029.g007

with (i) heat shock 27 kDa protein 1 (HSPB1), (ii) v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (KRAS), (iii) thymosin beta 4 X-linked (TMSB4X), and (iv) diacylglycerol kinase delta 130 kDa (DGKD). The annotations of these four interacting partners on their corresponding interaction pages on Wiki-Pi show that HSPB1 is involved in *blood vessel endothelial cell migration* and the other three, namely KRAS, TMSB4X, and DGKD, are all involved in *blood coagulation*. Further, KRAS annotations show that it is involved in *insulin receptor signaling pathway* (GO biological process). Researching for these genes outside of Wiki-Pi (i.e. in PubMed), it is also found that (i) TMSB4X may play a role in diabetic retinal neovascularization in the context of proliferative diabetic retinopathy [30], and that (ii) DGKD deficiency causes peripheral insulin resistance and metabolic inflexibility [31]. We conclude that IGSF21 may be involved in signaling cellular response to external stimuli, specifically triggering cytoskeletal organization and regulation of metabolic and catalytic activity, and that its association to diabetic retinopathy may be mediated through its interactions with the genes HSPB1, KRAS, TMSB4X and DGKD which are involved in blood-coagulation.

Conclusions

Wiki-Pi provides a means for effectively retrieving and studying human protein-protein interactions. The data itself is not curated by us, but is retrieved from other widely-used human protein information databases (Table 1). Wiki-Pi presents this information in a manner that is easy to be found and assimilated by biologists. The database is also timely because in the last few years several genome-wide association studies have been completed which resulted in the identification of genes associated with specific diseases or traits. Biological role of many of these genes is currently unknown or not fully characterized. If any such gene has known PPIs, the biological role of the gene may be determined based on the functions of its interacting partners.

Wiki-Pi facilitates the discovery of the molecular interconnects, if any, between seemingly unrelated biological processes that govern the human body: e.g. psychological stress and inflammation [32,33,34,35,36,37,38], or schizophrenia and immunity [32,39,40,41]; although these processes are hypothesized to be related, the molecular pathways connecting these processes are not well understood. Wiki-Pi makes it possible to search for interactions connecting these processes.

Biologists routinely draw inferences by putting together the information about the proteins and formulate hypotheses and conduct experiments to validate them; Wiki-Pi makes assimilation

of such information extremely easy by presenting all or most of the required annotations readily at hand. Wiki-Pi complements traditional databases, promoting research in molecular biology and biomedical informatics of human proteins. Future developments include the integration of additional data sources (both interactions and annotations) and the addition of authorship tracking for the wiki.

Supporting Information

File S1 Statistically enriched Gene Ontology biological process terms of PPIs of IGSF21. This figure is generated similar to Figure 6, but the node labels are shown and the image is in high resolution. Statistical significance of the node (GO term) is

shown in color, with the darker color indicating stronger significance. (PDF)

Acknowledgments

MG thanks Robert Sweet and Saumendra Sarkar for valuable discussions during the design of the website. Authors thank Adam Handen (undergraduate student at Rochester Institute of Technology) for creating the core components of the backend database, and Emily Cocco (High School student) for studying website usability.

Author Contributions

Conceived and designed the experiments: MG. Analyzed the data: NO. MG. Wrote the paper: MG NO. Developed software: NO.

References

- Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, et al. (2004) UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res* 32: D115–119.
- Prasad TS, Kandasamy K, Pandey A (2009) Human Protein Reference Database and Human Proteinpedia as discovery tools for systems biology. *Methods Mol Biol* 577: 67–79.
- Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D (1998) GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support. *Bioinformatics* 14: 656–664.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25: 25–29.
- Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28: 27–30.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28: 235–242.
- Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* 33: D514–517.
- Vastrik I, D'Eustachio P, Schmidt E, Gopinath G, Croft D, et al. (2007) Reactome: a knowledge base of biologic pathways and processes. *Genome Biol* 8: R39.
- Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, et al. (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* 34: D535–539.
- von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, et al. (2005) STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res* 33: D433–437.
- Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, et al. (2002) DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res* 30: 303–305.
- Chatr-aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, et al. (2007) MINT: the Molecular INTeraction database. *Nucleic Acids Res* 35: D572–574.
- Korb M, Rust AG, Thorsson V, Battail C, Li B, et al. (2008) The Innate Immune Database (IIDB). *BMC Immunol* 9: 7.
- Hermjakob H, Montecchi-Palazzi L, Lewington C, Mudali S, Kerrien S, et al. (2004) IntAct: an open source molecular interaction database. *Nucleic Acids Res* 32: D452–455.
- Hoffmann R (2008) A wiki for the life sciences where authorship matters. *Nat Genet* 40: 1047–1051.
- Harrison PJ, Weinberger DR (2005) Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. *Mol Psychiatry* 10: 40–68; image 45.
- Millar JK, Christie S, Anderson S, Lawson D, Hsiao-Wei Loh D, et al. (2001) Genomic structure and localisation within a linkage hotspot of Disrupted In Schizophrenia 1, a gene disrupted by a translocation segregating with schizophrenia. *Mol Psychiatry* 6: 173–178.
- Camargo LM, Collura V, Rain JC, Mizuguchi K, Hermjakob H, et al. (2007) Disrupted in Schizophrenia 1 Interactome: evidence for the close connectivity of risk genes and a potential synaptic basis for schizophrenia. *Mol Psychiatry* 12: 74–86.
- Wang Q, Jaaro-Peled H, Sawa A, Brandon NJ (2008) How has DISC1 enabled drug discovery? *Mol Cell Neurosci* 37: 187–195.
- Camargo LM, Wang Q, Brandon NJ (2008) What can we learn from the disrupted in schizophrenia 1 interactome: lessons for target identification and disease biology? *Novartis Found Symp* 289: 208–216; discussion 216–221, 238–240.
- Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, et al. (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res* 36: D901–906.
- Maglott D, Ostell J, Pruitt KD, Tatusova T (2011) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res* 39: D52–57.
- Hubbard T, Barker D, Birney E, Cameron G, Chen Y, et al. (2002) The Ensembl genome database project. *Nucleic Acids Res* 30: 38–41.
- Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21: 3448–3449.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498–2504.
- Baran J, Gerner M, Haeussler M, Nenadic G, Bergman CM (2011) pubmed2ensembl: A Resource for Mining the Biological Literature on Genes. *PLoS One* 6: e24716.
- Manning CD, Raghavan P, Schütze H (2008) Introduction to information retrieval. New York: Cambridge University Press. xxi, 482 p.p.
- Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, et al. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 106: 9362–9367.
- Grassi MA, Tikhomirov A, Ramalingam S, Below JE, Cox NJ, et al. (2011) Genome-wide meta-analysis for severe diabetic retinopathy. *Hum Mol Genet* 20: 2472–2481.
- Wang JY, Lu Q, Tao Y, Jiang YR, Jonas JB (2011) Intraocular expression of thymosin beta4 in proliferative diabetic retinopathy. *Acta Ophthalmol* 89: e396–403.
- Chibalin AV, Leng Y, Vieira E, Krook A, Bjornholm M, et al. (2008) Downregulation of diacylglycerol kinase delta contributes to hyperglycemia-induced insulin resistance. *Cell* 132: 375–386.
- Aרון D, Unger T, Lewis DA, Levitt P, Mirmics K (2007) Molecular evidence for increased expression of genes related to immune and chaperone function in the prefrontal cortex in schizophrenia. *Biol Psychiatry* 62: 711–721.
- Jia P, Wang L, Meltzer HY, Zhao Z (2010) Common variants conferring risk of schizophrenia: a pathway analysis of GWAS data. *Schizophr Res* 122: 38–42.
- Esch T, Stefano GB, Fricchione GL, Benson H (2002) The role of stress in neurodegenerative diseases and mental disorders. *Neuro Endocrinol Lett* 23: 199–208.
- Gold SM, Irwin MR (2009) Depression and immunity: inflammation and depressive symptoms in multiple sclerosis. *Immunol Allergy Clin North Am* 29: 309–320.
- Leonard BE, Myint A (2009) The psychoneuroimmunology of depression. *Hum Psychopharmacol* 24: 165–175.
- Lutgendorf SK (2009) Positive affect and radiation-induced inflammation: insights into inflammatory regulation? *Brain Behav Immun* 23: 1066–1067.
- Skinner R, Georgiou R, Thornton P, Rothwell N (2009) Psychoneuroimmunology of stroke. *Immunol Allergy Clin North Am* 29: 359–379.
- Brown AS, Derkits EJ (2010) Prenatal infection and schizophrenia: a review of epidemiologic and translational studies. *Am J Psychiatry* 167: 261–280.
- Jabocs K (2006) Schizophrenia and the immune system. *Harvard Science Review Spring* 2006: 46–48.
- Crack PJ, Bray PJ (2007) Toll-like receptors in the brain and their potential roles in neuropathology. *Immunol Cell Biol* 85: 476–480.