Biotechnological Communication

Biosci. Biotech. Res. Comm. 10(2): 13-18 (2017)

BBRC Bioscience Biotechnology Research Communications

Edifice of vitiDB: A first ever structured portal for vitiligo protein repository and its assessment

Anvita Gupta Malhotra, Sudha Singh, Mohit Jha and Khushhali M Pandey*

Department of Biological sciences and Engineering, MANIT, Bhopal (M.P) INDIA

ABSTRACT

vitiDB (http://vitidb.com/index.php) is the first ever structured repository exclusively for vitiligo disorder. This portal brings together valuable but heterogeneous or disparate information available from a spectrum of public domain databases, patents and published literature sources on vitiligo. The aim is to arm the research community with critical details extracted froma reviewed resource of vitiligo genes and proteins. In addition, the database throws light on interectome of proteins and their targetability assessment. The current release of vitiDB contains 333 vitiligo disease - associated proteins and nearly 5000 allied proteins. Each entry provides comprehensive data related to the protein like its kinetic, pharmacological and ontological properties. Additionally, this portal provides browsing and extracting information related to vitiligo protein interaction network and its topological properties. vitiDB catalogues 107416 unique interactions among 4845 proteins derived from 8 different databases. Detailed targetability analysis is available at this portaland it includes druggability, assayability, essentiality, vulnerability and secretability analysis for the disease proteins. This user-friendly web interface will help the user to have an informed opinion on disease genes without having to plough through various databases. The additional information in the form of derived data can potentially assist in the drug discovery process. There will be sustained effortsaimed at periodic updation of the core data in the wake of their extension / modification as well as constructive feedback received from the users.

KEY WORDS: VITILIGO, INTERACTION NETWORK, TARGETABILITY, DRUGGABILITY, DATABASE

ARTICLE INFORMATION:

**Corresponding Author: kmpbiomanit@gmail.com, khushhalimenariya@manit.ac.in*

Received 27th Nov, 2016
Accepted after revision 26th Jan, 2017
BBRC Print ISSN: 0974-6455
Online ISSN: 2321-4007 CODEN: USA BBRCBA
Thomson Reuters ISI ESC and Crossref Indexed Journal
NAAS Journal Score 2017: 4.31 Cosmos IF : 4.006

Online Contents Available at: http//www.bbrc.in/

Introduction

Vitiligo is a progressive depigmentation disorder affecting nearly 1% of the world population (Guerra et al., 2010). Vitiligo disorder is characterized by progressive skin depigmentation phenomena, which is induced and maintained by the loss of melanin - producing cells at the cutaneous level. Vitiligo is known to be a skin disorder with complex etiology. Its pathogenesis and the inner causes are still unclear but, in recent years, scientific research identified various pathways and processes for its onset and progression. There are several proposed hypotheses for this depigmenting disorder which includes autoimmunity auto-cytotoxic/ metabolic mechanisms, (Reimann et al., 2012), and impaired melanocyte migration and/or proliferation (Westerhof and d'Ischia, 2007, Speeckaert et al., 2015).

Studies also show the important role played by genetic susceptibility in the origin of vitiligo. These mechanisms are neither mutually exclusive, noradequate enough to explain the disease etiology on their own. An integrated viewpoint can be formed by incorporating all the different causal factors that could contribute to some extent to melanocyte destruction into a 'convergence theory' (Le Poole et al., 1993), (Schallreuter et al., 2008, Jin et al., 2016, Spritz, 2011). Vitiligo involves a complex framework of interactions between various proteins, pathways and processes (Laddha et al., 2013). An attempt has been made here to bring together these disease proteins on a common platform to display their related properties. Further protein-protein interaction (PPI) data for each of these disease proteins were extracted to generate a comprehensive vitiligo interectome map (Malhotra et al., 2017).

Information regarding the interectome analysis is also made readily available on this portal.A better understanding of the disease could provide new targets for prevention or treatment of vitiligo (Passeron and Ortonne, 2012). An ideal drug target should comprehend following properties: favourable assayability for high throughput screening, ability to modify a disease, less effect on alteration of physiological conditions or other diseases, differential expression across the body, the availability of a biomarker, etc. Experimentally evaluating all proteins for their targetability is an overwhelming task (Gashaw et al., 2011, Dey-Rao and Sinha, 2017).

Hence, it's not experimentally possible to analyse and prioritize all the drug targets in a laboratory. This necessitates employinga computational technique for evaluation of proteins on the standards of potent drug targets (Sakharkar and Sakharkar, 2007, Costa et al., 2010, Kandoi et al., 2015). Thus targetability analysis of these proteins will aid in shortlisting prospective therapeutic drug targetsfrom the entire proteome with high specificity.

There is no exclusive database for vitiligo disease proteins available till date. All related information is dispersed in various published literature and online databases and is, therefore, extremely difficult to access. For assisting the vitiligo researchers, all scattered data has been compiled together and adedicated and comprehensive vitiligo proteins repository – vitiDB - is developed.

Methodology

Data collection and compilation

The organization of data was done systematically as primary and derived data [Figure-1]. The primary data consists of the vitiligo disease proteins related information. The secondary data includes disease proteins interaction data and its analysis features. Alongside the targetability related parameters were also computed and added to the resource.

Disease Protein Information

In order to build a comprehensive resource for vitiligo proteins an extensive search was carried out across various online databases like NCBI (2017), Uniprot (2015), AnyGene [AnyGenes® http://www.anygenes.com/index.php], Harmonizome (Rouillard et al., 2016), vitivar [http://vitivar.igib.res.in/genes], GeneCards (Rebhan et al., 1998), DOlite (Du et al., 2009), DisGeNET (Pinero et al., 2015), OMIM (Hamosh et al., 2005), Disease Ontology (Schriml et al., 2012) and Human Phenotype Ontology (Kohler et al., 2014). Along with these databases, information was also manually curated from published literature, microarray experiments and SNP analysis. Gene susceptibility studies (Passeron and Ortonne, 2005), (Picardo et al., 2015) were also considered while collating the disease genes for vitiligo.

Vitiligo Interectome Data

PPI data for the disease proteins were extracted from eight different databases. These includes BioGrid (Chatr-Aryamontri et al., 2015), MINT (Ceol et al., 2010), STRING (Szklarczyk et al., 2017), Reactome (Croft et al., 2014), DIP (Xenarios et al., 2002), IntAct (Kerrien et al., 2012), Spike (Paz et al., 2011) and GeneMania (Zuberi et al., 2013). All this data was merged and the unique interactions were employed for construction of unidirectional vitiligo interaction map (Malhotra et al., 2017) in Cytoscape (Kohl et al., 2011).

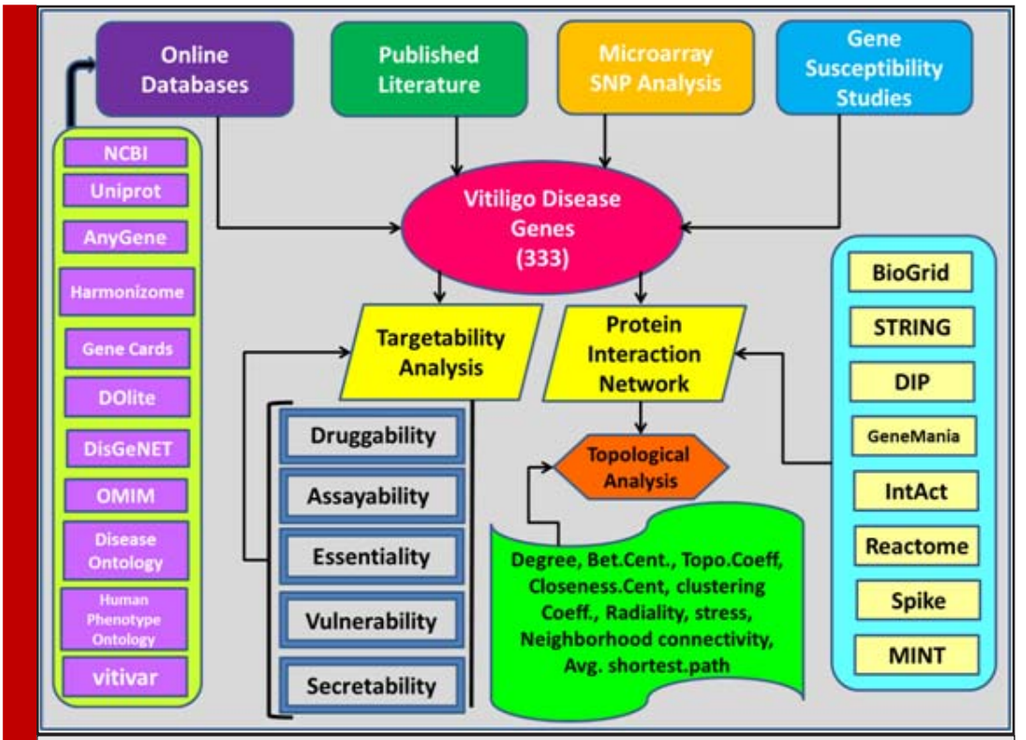

FIGURE 1. Architecture of vitiDB.

Cytoscape is a highly versatile program for the analysis, operation and visualization of large networks. Later, the topological properties (Degree, Betweeness centrality, Topological coefficient, Clustering coefficient, Closeness centrality, Radiality, Stress, Neighbourhood connectivity and Average shortest path length) were calculated for all the proteins in the map. These calculation were done in Network Analyzer (Assenov et al., 2008), an inbuilt plugin with cytoscape that is used for comprehensive analysis of network topologies.

Targetability statistics for disease proteins

The key disease proteins were analysed for their ability to be a good therapeutic target (Rask-Andersen et al., 2011). This was done under five heads namely Druggability (Druggable Genome (Hopkins and Groom, 2002), can SAR druggability (Bulusu et al., 2014), DrugBank (Law et al., 2014), DGI druggability (Griffith et al., 2013), PDTD (Gao et al., 2008) and Domain DrugEBIlity (EMBL) [https://www.ebi.ac.uk/chembl/drugebility/]) Assayability (Sigma Aldrich database, Brenda (Schomburg et al., 2017), Ki database [https://kidbdev.med.unc.edu/databases/kidb.php], PDB (Berman et al., 2000) or Binding DB (Gilson et al., 2016)), Essentiality (OGE (Chen et al., 2012), MGD (Eppig et al., 2007), Human Phenotype project (Kohler et al., 2014), Part et al. (Park et al., 2008) or Georgi et al. (Georgi et al., 2013)), Vulnerability (Holme et al., 2002) and Secretability (Narayanan, 2015).

Database architecture and web interface development

All the collated data were entered in excel files which were converted into csv files. These files were later imported into MySQL database. MySQL, an object-relational open source database management system, was employed to manage the data at the back-end. There were 11 tables in the database. The database was launched using Apache HTTP server online on Linux Platform and on local machine via WAMP server 2.0. For the backend support i.e., database interfacing scripts, we have used PHP programming language. It queries backend database to retrieve information. The front end was designed using HTML 5.

Results and Discussion

Implementation

The database is sectioned into three categories: Vitiligo Protein Repository, Vitiligo Interectome map and Targetability Analysis. All the information can be browsed under the sections mentioned below (Figure-2).

FIGURE 2. Schematic representation of vitiDB web interface.

Vitiligo Protein Repository:

A total of 333 disease proteins were identified. Their related information is made available from this section under the following features:

Basic Information- This includes Gene name, Protein name, Uniprot id, status, Length of the protein, gene name synonyms and also the source from which this protein is obtained as vitiligo disease protein. This source can be related to protein or pathway in which the protein is involved like database, microarray experimentation, autoimmunity, susceptibility studies, apoptosis, melanogenesis (Singh et al., 2013) or oxidative stress.

Kinetic properties – Enzyme related information is displayed here under the headings EC number, Catalytic activity, KineticsPathway, Enzyme regulation, Active site, Binding site, Site or Function.

Ontology Data – Gene Ontology details and identifiers are made available.

PubmedDetails – Pubmed ids for the corresponding protein is mentioned.

Cross-reference to other online databases - The vitiligo protein entries are linked to other biological databases like PDB, SMR, Protein Model Portal, PDBsum, DisProt, STRING, MINT, IntAct, DIP, BioGrid, Guide to PHARMACOLOGY, DrugBank, ChEMBL, BindingDB, KEGG, GeneID, Ensembl, BRENDA, BioCyc, Reactome, InterPro, PRINTS, PROSITE, Pfam, DisGeNET, GeneCards, PharmGKB or GeneWiki.

Vitiligo Interaction Map

The disease proteins interaction network consists of 4845 nodes and 107416 unique interactions. The interacting partners along with the topological properties are available for each of 4845 proteins. Further, based on the network topology analysis, the proteins were classified as backbone and core proteins and are labeled accordingly in the database under the section *Classification of proteins.*

Targetability Analysis of Disease proteins

A comprehensive description of all the features pertaining to the drug target identification was fetched from various sources and displayed under this head. The proteins that are in alignment with most of the parameters are entitledto be the prospective target for the vitiligo disorder. The later updates will aim at making the targetability analysis available for all the human proteins.

Concluding Remarks

The increasing amount of data on protein interactions, drug target features,and the knowledge of drug-like properties can act as a very lucrative starting point for vitiligo drug hunt. Vitiligo sufferers still have to face ostracism and discrimination around the world. This is a modest attempt to aid in its treatment. The enriched data and search utilities available at vitiDB might assist other

online databases in providing all-inclusive information about the drugs, targets and target identification process aimed at further research and drug discovery efforts.

Availability

vitiDBis a free database that can be accessed at http://vitidb.com

Acknowledgments

The authors thankfully acknowledge the help, support and guidance provided by Dr. Ajay Pandey, Department of Mechanical Engineering, MANIT, Bhopal.

References

Anonymous (2015) UniProt: a hub for protein information, Nucleic Acids Res 43(Database issue), D204-212.

Anonymous (2017) Database Resources of the National Center for Biotechnology Information, Nucleic Acids Res 45(D1), D12-D17.

Assenov, Y., F. Ramirez, S. E. Schelhorn, T. Lengauer, M. Albrecht (2008), Computing topological parameters of biological networks, Bioinformatics 24(2), 282-284.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne (2000) The Protein Data Bank, Nucleic Acids Res 28(1), 235-242.

Bulusu, K. C., J. E. Tym, E. A. Coker, A. C. Schierz, B. Al-Lazikani (2014) canSAR: updated cancer research and drug discovery knowledgebase, Nucleic Acids Res 42(Database issue), D1040- 1047.

Ceol, A., A. ChatrAryamontri, L. Licata, D. Peluso, L. Briganti, L. Perfetto, L. Castagnoli, G. Cesareni (2010) MINT: the molecular interaction database, 2009 update, Nucleic Acids Res 38(Database issue), D532-539.

Chatr-Aryamontri, A., B. J. Breitkreutz, R. Oughtred, et al., (2015) The BioGRID interaction database: 2015 update, Nucleic Acids Res 43(Database issue), D470-478.

Chen, W. H., P. Minguez, M. J. Lercher, P. Bork (2012) OGEE: an online gene essentiality database, Nucleic Acids Res 40(Database issue), D901-906.

Costa, P. R., M. L. Acencio, N. Lemke (2010) A machine learning approach for genome-wide prediction of morbid and druggable human genes based on systems-level data, BMC Genomics 11 Suppl 5, S9.

Croft, D., A. F. Mundo, R. Haw, M. Milacic, et al., (2014) The Reactome pathway knowledgebase, Nucleic Acids Res 42(Database issue), D472-477.

Dey-Rao, R., A. A. Sinha (2017) Vitiligo blood transcriptomics provides new insights into disease mechanisms and identifies potential novel therapeutic targets, BMC Genomics 18(1), 109.

Du, P., G. Feng, J. Flatow, J. Song, M. Holko, W. A. Kibbe, S. M. Lin (2009) From disease ontology to disease-ontology lite, statistical methods to adapt a general-purpose ontology for the test of gene-ontology associations, Bioinformatics 25(12), i63-68.

Eppig, J. T., J. A. Blake, C. J. Bult, J. A. Kadin, J. E. Richardson (2007) The mouse genome database (MGD): new features facilitating a model system, Nucleic Acids Res 35(Database issue), D630-637.

Gao, Z., H. Li, H. Zhang, X. Liu, L. Kang, X. Luo, W. Zhu, K. Chen, X. Wang, H. Jiang (2008) PDTD: a web-accessible protein database for drug target identification, BMC Bioinformatics 9, 104.

Gashaw, I., P. Ellinghaus, A. Sommer, K. Asadullah (2011) What makes a good drug target?, Drug Discov Today 16(23-24), 1037-1043.

Georgi, B., B. F. Voight, M. Bucan (2013) From mouse to human, evolutionary genomics analysis of human orthologs of essential genes, PLoS Genet 9(5), e1003484.

Gilson, M. K., T. Liu, M. Baitaluk, G. Nicola, L. Hwang, J. Chong (2016) BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology, Nucleic Acids Res 44(Database issue), D1045-1053.

Griffith, M., O. L. Griffith, A. C. Coffman, J. V. Weible, J. F. McMichael, N. C. Spies, J. Koval, I. Das, M. B. Callaway, J. M. Eldred, C. A. Miller, J. Subramanian, R. Govindan, R. D. Kumar, R. Bose, L. Ding, J. R. Walker, D. E. Larson, D. J. Dooling, S. M. Smith, T. J. Ley, E. R. Mardis, R. K. Wilson (2013) DGIdb - Mining the druggable genome, Nat Methods 10(12), 1209-1210.

Guerra, L., E. Dellambra, S. Brescia, D. Raskovic (2010) Vitiligo: pathogenetic hypotheses and targets for current therapies, Curr Drug Metab 11(5), 451-467.

Hamosh, A., A. F. Scott, J. S. Amberger, C. A. Bocchini, V. A. McKusick (2005) Online Mendelian Inheritance in Man (OMIM): a knowledgebase of human genes and genetic disorders, Nucleic Acids Res 33(Database issue), D514-517.

Holme, P., B. J. Kim, C. N. Yoon, S. K. Han (2002) Attack vulnerability of complex networks, Phys Rev E Stat Nonlin Soft Matter Phys 65(5 Pt 2), 056109.

Hopkins, A. L., C. R. Groom (2002) Thedruggable genome, Nat Rev Drug Discov 1(9), 727-730.

Jin, Y., G. Andersen, D. Yorgov, T. M. Ferrara, S. Ben, K. M. Brownson, P. J. Holland, S. A. Birlea, J. Siebert et al., (2016) Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants, Nat Genet 48(11), 1418-1424.

Kandoi, G., M. L. Acencio, N. Lemke (2015) Prediction of Druggable Proteins Using Machine Learning and Systems Biology, A Mini-Review, Front Physiol 6.

Kerrien, S., B. Aranda, L. Breuza, A. Bridge, et al., (2012) The IntAct molecular interaction database in 2012, Nucleic Acids Res 40(Database issue), D841-846.

Kohl, M., S. Wiese, B. Warscheid (2011) Cytoscape: software for visualization and analysis of biological networks, Methods MolBiol 696, 291-303.

Kohler, S., S. C. Doelken, C. J. Mungall, S. Baueret al., (2014) The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data, Nucleic Acids Res 42(Database issue), D966-974.

Laddha, N. C., M. Dwivedi, M. S. Mansuri, A. R. Gani, M. et al (2013) Vitiligo: interplay between oxidative stress and immune system, ExpDermatol 22(4), 245-250.

Law, V., C. Knox, Y. Djoumbou, T. Jewison, et al., (2014) DrugBank 4.0: shedding new light on drug metabolism, Nucleic Acids Res 42(Database issue), D1091-1097.

Le Poole, I. C., P. K. Das, R. M. van den Wijngaard, J. D. Bos, W. Westerhof (1993) Review of the etiopathomechanism of vitiligo: a convergence theory, ExpDermatol 2(4), 145-153.

Malhotra, A. G., M. Jha, S. Singh, K. M. Pandey (2017) Construction of a Comprehensive Protein-Protein Interaction Map for Vitiligo Disease to Identify Key Regulatory Elements: A Systemic Approach, Interdiscip Sci.

Narayanan, R. (2015) Druggable Vitiligo Genome, A Fast Track Approach to Take the Genome Wide Association to the Clinic, MOJ Proteomics & Bioinformatics 2(3).

Park, D., J. Park, S. G. Park, T. Park, S. S. Choi (2008) Analysis of human disease genes in the context of gene essentiality, Genomics 92(6), 414-418.

Passeron, T., J. P. Ortonne (2005) Physiopathology and genetics of vitiligo, J Autoimmun 25 Suppl, 63-68.

Passeron, T., J. P. Ortonne (2012) Activation of the unfolded protein response in vitiligo: the missing link?" J Invest Dermatol 132(11), 2502-2504.

Paz, A., Z. Brownstein, Y. Ber, S. Bialik, E. et al., (2011) SPIKE: a database of highly curated human signaling pathways, Nucleic Acids Res 39(Database issue), D793-799.

Picardo, M., M. L. Dell'Anna, K. Ezzedine, I. Hamzavi, J. E. Harris, D. Parsad, A. Taieb (2015) Vitiligo, Nat Rev Dis Primers 1, 15011.

Pinero, J., N. Queralt-Rosinach, A. Bravo, J. Deu-Pons, A. Bauer-Mehren, M. Baron, F. Sanz, L. I. Furlong (2015) DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes, Database (Oxford) 2015, bav028.

Rask-Andersen, M., M. S. Almen, H. B. Schioth (2011) Trends in the exploitation of novel drug targets, Nat Rev Drug Discov 10(8), 579-590.

Rebhan, M., V. Chalifa-Caspi, J. Prilusky, D. Lancet (1998) GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support, Bioinformatics 14(8), 656-664.

Reimann, E., K. Kingo, M. Karelson, P. Reemann, U. Loite, M. Keermann, K. Abram, E. Vasar, H. Silm, S. Koks (2012) Expression profile of genes associated with the dopamine pathway in vitiligo skin biopsies and blood sera, Dermatology 224(2), 168-176.

Rouillard, A. D., G. W. Gundersen, N. F. Fernandez, Z. Wang, C. D. Monteiro, M. G. McDermott, A. Ma'ayan (2016) The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins, Database (Oxford) 2016.

Sakharkar, M. K., K. R. Sakharkar (2007) Targetability of human disease genes, Curr Drug Discov Technol 4(1), 48-58.

Schallreuter, K. U., P. Bahadoran, M. Picardo, et al., (2008) Vitiligo pathogenesis: autoimmune disease, genetic defect, excessive reactive oxygen species, calcium imbalance, or what else?" ExpDermatol 17(2), 139-140; discussion 141-160.

Schomburg, I., L. Jeske, M. Ulbrich, S. Placzek, A. Chang, D. Schomburg (2017) The BRENDA enzyme information system-From a database to an expert system, J Biotechnol.

Schriml, L. M., C. Arze, S. Nadendla, Y. W. Chang, M. Mazaitis, V. Felix, G. Feng, W. A. Kibbe (2012) Disease Ontology: a backbone for disease semantic integration, Nucleic Acids Res 40(Database issue), D940-946.

Singh, S., A. G. Malhotra, A. Pandey, K. M. Pandey (2013) Computational model for pathway reconstruction to unravel the evolutionary significance of melanin synthesis, Bioinformation 9(2), 94.

Speeckaert, R., M. M. Speeckaert, N. van Geel (2015) Why treatments do(n't) work in vitiligo: An autoinflammatory perspective, Autoimmun Rev 14(4), 332-340.

Spritz, R. A. (2011) Recent progress in the genetics of generalized vitiligo, J Genet Genomics 38(7), 271-278.

Szklarczyk, D., J. H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N. T. Doncheva, A. Roth, P. Bork, L. J. Jensen, C. von Mering (2017) The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible, Nucleic Acids Res 45(Database issue), D362-368.

Westerhof, W., M. d'Ischia (2007) Vitiligo puzzle, the pieces fall in place, Pigment Cell Res 20(5), 345-359.

Xenarios, I., L. Salwinski, X. J. Duan, P. Higney, S. M. Kim, D. Eisenberg (2002) DIP: the Database of Interacting Proteins, a research tool for studying cellular networks of protein interactions, Nucleic Acids Res 30(1), 303-305.

Zuberi, K., M. Franz, H. Rodriguez, J. Montojo, C. T. Lopes, G. D. Bader, Q. Morris (2013) GeneMANIA prediction server 2013 update, Nucleic Acids Res 41(Web Server issue), W115-122.