Chapter 7

Conclusion

7.1 Concluding Remarks

The enormous growth in biological data has paved way for the computer scientists to intervene in their analysis and infer useful information in this area of computational problems. The understanding of the association of the different phenotypic traits with that of genes or proteins is increasing day by day with the combined efforts of the biologists and computer scientists. However, there still lies a huge gap in associating the findings from the technological grounds with that from the biologists' point of view. I have used certain data mining techniques to propose solutions to few such computational problems that lie on the pathway from genes to proteins and finally their association with the phenotypic traits within an individual. I have used the Protein Protein Interaction data (PPI) and the gene expression data in my work.

The first part of my research deals with grouping proteins obtained from raw PPI data into clusters known as protein complexes. The obtained complexes need to be of high biological significance. This part of the work is then extended towards inferencing disease related information such as the topological association of disease genes with other genes in the network, their role in the network etc. Another work done in extension with the PPI data is the ranking of disease associated complexes. The second part of my research work deals with analyzing the expression changes occurring in genes as a disease progresses from one stage to another. This analysis is carried out based on the similarities that exist between genes in a cluster w.r.t. their status, i.e., to which stage of the disease they belong to. This work has also been extended towards identifying novel biomarkers which can be associated directly or indirectly with the disease under consideration. Now, I report my observations in

the analysis of PPI and gene expression data during the course of my PhD work.

- The first protein complex finding method, CNCM which is reported in Chapter 3 is capable of performing consistently well across a range of datasets with minimum parameter tuning. This is one of the most important requirement for any good complex finding method. The complexes obtained using this method have been shown to be of high biological significance w.r.t. the benchmark datasets.

- The next part of Chapter 3 discusses a method called DCRS which is based on the multi-objective optimization technique. It uses the preprocessed data obtained from CNCM as its initial population and gives a set of complexes of high biological relevance. This method has been validated over human PPI dataset.

- Chapter 4 reports a method called CSC for protein complex finding. This method uses a combination of both topological and biological property of the PPI network in order to detect quality complexes. It has been validated over both yeast and human PPI dataset. This work is extended towards identifying the association of Alzheimer's disease genes in complexes, their position w.r.t. other genes and also in determining the links which connects them with other genes of the network.

- A method called ComFiR is also reported in Chapter 4 for complex detection from PPI networks. It is based on a semi-supervised approach and uses information from benchmark complexes to extract further quality complexes. The results obtained by this method has been shown to perform better than all other existing methods in case of human PPI dataset. An effective ranking scheme has been introduced to rank the complexes. The ranking scheme has been shown for Alzheimer's Disease.

- Chapter 5 reports a gene module extraction technique based on the semantic similarity between genes. This method has been used over a two-stage breast cancer dataset. Novel biomarkers associated with the progression of the disease are identified. These biomarkers have been established from gene expression, pathway and literature sources.

- A subspace based gene-gene network construction and a module extraction technique has been proposed in Chapter 6. This method has been used over Parkinson's Disease dataset. Some interesting biomarkers associated with the

disease are reported in this work. These biomarkers have been established using literature sources. An analysis of the hub genes found in modules is also reported here. This analysis is based on a number of centrality measures.

7.2 Future Works

While dealing with the issues involved in PPI and gene expression data and its association with various physiological processes, a number of research directions were realized. These are as follows.

- PPI data and gene expression data were used substantially for analyzing the physiological characteristics of living organisms. Apart from these data, miRNA and RNA-Seq data can also be used for effectively analyzing the properties of genes and their respective roles.
- Proper use of biological properties of PPI network can lead to improved complex finding results.
- All the analysis done in this research work is based on unweighted networks, analyzing the proposed methods in terms of weighted networks may produce significant results.
- The subspace network construction technique proposed here is very time consuming. Using a parallel environment for such computation would ease the whole process.
- An integrated framework for protein complex finding and its validation is one of the immediate concerns in the area of PPI data analysis.
- A combination of a number of available biological data such as gene expression, PPI data, RNA-Seq data in addition to clinical data could be used for effective analysis of genes and their behavioral changes in case of diseases.

Bibliography

[1] Ahmed, H. A., Mahanta, P., Bhattacharyya, D. K., and Kalita, J. K. Module extraction from subspace co-expression networks. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 1(4):183–195, 2012.

[2] Alizadeh, A. A., Elsen, M. B., Davis, R. E., Ma, C., et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, 403(6769):503, 2000.

[3] Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., et al. Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25(1):25, 2000.

[4] Bader, G. D. and Hogue, C. W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*, 4(1): 2, 2003.

[5] Bandyopadhyay, S., Ray, S., Mukhopadhyay, A., and Maulik, U. A multi-objective approach for identifying protein complexes and studying their association in multiple disorders. *Algorithms for Molecular Biology*, 10(1):24, 2015.

[6] Bar-Joseph, Z. Analyzing time series gene expression data. *Bioinformatics*, 20(16):2493–2503, 2004.

[7] Barkow, S., Bleuler, S., Prelić, A., Zimmermann, P., and Zitzler, E. Bicat: a biclustering analysis toolbox. *Bioinformatics*, 22(10):1282–1283, 2006.

[8] Barrero, F., Ampuero, I., Morales, B., Vives, F., Del Castillo, J. d. D. L., Hoenicka, J., and Yebenes, J. G. Depression in Parkinson's disease is related to a genetic polymorphism of the cannabinoid receptor gene (CNR1). *The Pharmacogenomics Journal*, 5(2):135, 2005.

[9] Barrett, T. and Edgar, R. Gene Expression Omnibus: Microarray Data Storage, Submission, Retrieval, and Analysis. *Methods in Enzymology*, 411: 352–369, 2006.

[10] Bass, J. I. F., Diallo, A., Nelson, J., Soto, J. M., Myers, C. L., and Walhout, A. J. Using networks to measure similarity between genes: association index selection. *Nature Methods*, 10(12):1169, 2013.

[11] Becker, E., Robisson, B., Chapple, C. E., Guénoche, A., and Brun, C. Multifunctional proteins revealed by overlapping clustering in protein interaction network. *Bioinformatics*, 28(1):84–90, 2012.

[12] Beer, D. G., Kardia, S. L., Chiang-Ching, H., Giordano, T. J., Levin, A. M., Misek, D. E., Lin, L., Chen, G., Gharib, T. G., Thomas, D. G., et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nature Medicine*, 8(8):816, 2002.

[13] Berriz, G. F., King, O. D., Bryant, B., Sander, C., and Roth, F. P. Characterizing gene sets with FuncAssociate. *Bioinformatics*, 19(18):2502–2504, 2003.

[14] Brazma, A. and Vilo, J. Gene expression data analysis. *FEBS Letters*, 480 (1):17–24, 2000.

[15] Chen, Y., Jacquemin, T., Zhang, S., and Jiang, R. Prioritizing protein complexes implicated in human diseases by network optimization. In *BMC Systems Biology*, volume 8, page S2. BioMed Central, 2014.

[16] Chu, Y. and Corey, D. R. Rna sequencing: platform selection, experimental design, and data interpretation. *Nucleic Acid Therapeutics*, 22(4):271–274, 2012.

[17] Chua, H. N., Ning, K., Sung, W.-K., Leong, H. W., and Wong, L. Using indirect protein–protein interactions for protein complex prediction. *Journal of Bioinformatics and Computational Biology*, 6(03):435–466, 2008.

[18] Clancy, S. and Brown, W. Translation: DNA to mRNA to protein. *Nature Education*, 1(1):101, 2008.

[19] Cohen, D., Chumakov, I., Nabirochkin, S., Guerassimenko, O., and Graudens, E. New diagnostic tools for alzheimer disease, July 29 2010. US Patent App. 13/387,174.

[20] Cook, A. L., Vitale, A. M., Ravishankar, S., Matigian, N., Sutherland, G. T., Shan, J., Sutharsan, R., Perry, C., Silburn, P. A., Mellick, G. D., et al. NRF2 activation restores disease related metabolic deficiencies in olfactory neurosphere-derived cells from patients with sporadic Parkinson's disease. *PLOS One*, 6(7):e21907, 2011.

[21] Cox, M. M. and Nelson, D. L. *Lehninger principles of biochemistry.* WH Freeman. ISBN ISBN-10: 1-4641-2611-9.

[22] Crick, F. Central dogma of molecular biology. *Nature*, 227(5258):561, 1970.

[23] Danford, T., Rolfe, A., and Gifford, D. GSE: a comprehensive database system for the representation, retrieval, and analysis of microarray data. In *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, page 539. NIH Public Access, 2008.

[24] Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.

[25] Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., and Lempicki, R. A. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biology*, 4(9):R60, 2003.

[26] Deshpande, V., Sharma, A., Mukhopadhyay, R., Thota, L. N. R., Ghatge, M., Vangala, R. K., Kakkar, V. V., and Mundkur, L. Understanding the progression of atherosclerosis through gene profiling and co-expression network analysis in Apob tm2sgy ldlr tm1Her double knockout mice. *Genomics*, 107 (6):239–247, 2016.

[27] Dranoff, G. Cytokines in cancer pathogenesis and cancer therapy. *Nature Reviews Cancer*, 4(1):11, 2004.

[28] Enright, A. J., Iliopoulos, I., Kyrpides, N. C., and Ouzounis, C. A. Protein interaction maps for complete genomes based on gene fusion events. *Nature*, 402(6757):86–90, 1999.

[29] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. From data mining to knowledge discovery in databases. *AI Magazine*, 1996.

[30] Ferrari, F., Bortoluzzi, S., Coppe, A., Sirota, A., Safran, M., Shmoish, M., Ferrari, S., Lancet, D., Danieli, G. A., and Bicciato, S. Novel definition files

for human GeneChips based on GeneAnnot. *BMC Bioinformatics*, 8(1):446, 2007.

[31] Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., Von Mering, C., et al. STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Research*, 41(D1):D808–D815, 2012.

[32] Freeman, W. M., Walker, S. J., and Vrana, K. E. Quantitative RT-PCR: pitfalls and potential. *Biotechniques*, 26:112–125, 1999.

[33] Gambette, P., Guénoche, A., et al. Bootstrap clustering for graph partitioning. *RAIRO-Operations Research*, 2011.

[34] Ganegoda, G. U., Wang, J., Wu, F.-X., and Li, M. Prediction of disease genes using tissue-specified gene-gene network. *BMC Systems Biology*, 8(Suppl 3): S3, 2014.

[35] Gavin, A.-C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A.-M., Cruciat, C.-M., et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, 415(6868):141–147, 2002.

[36] Gavin, A.-C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dümpelfeld, B., et al. Proteome survey reveals modularity of the yeast cell machinery. *Nature*, 440(7084):631–636, 2006.

[37] Ghanipour, A., Jirström, K., Pontén, F., Glimelius, B., Påhlman, L., and Birgisson, H. The prognostic significance of tryptophanyl-tRNA synthetase in colorectal cancer. *Cancer Epidemiology and Prevention Biomarkers*, 18 (11):2949–2956, 2009.

[38] Girault, J.-A. and Greengard, P. The neurobiology of dopamine signaling. *Archives of Neurology*, 61(5):641–644, 2004.

[39] Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, 286(5439):531–537, 1999.

[40] Guénoche, A. Consensus of partitions: a constructive approach. *Advances in Data Analysis and Classification*, 5(3):215–229, 2011.

[41] Hahn, M. W. and Kern, A. D. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Molecular Biology and Evolution*, 22(4):803–806, 2005.

[42] Hamza, T. H., Chen, H., Hill-Burns, E. M., Rhodes, S. L., Montimurro, J., Kay, D. M., Tenesa, A., Kusel, V. I., Sheehan, P., Eaaswarkhanth, M., et al. Genome-wide gene-environment study identifies glutamate receptor gene GRIN2A as a Parkinson's disease modifier gene via interaction with coffee. *PLOS Genetics*, 7(8):e1002237, 2011.

[43] Han, J., Pei, J., and Kamber, M. *Data Mining: Concepts and Techniques*. Elsevier, 2011.

[44] Hartwell, L. H., Hopfield, J. J., Leibler, S., and Murray, A. W. From molecular to modular cell biology. *Nature*, 402(6761supp):C47, 1999.

[45] He, D., Liu, Z.-P., Honda, M., Kaneko, S., and Chen, L. Coexpression network analysis in chronic hepatitis B and C hepatic lesions reveals distinct patterns of disease progression to hepatocellular carcinoma. *Journal of Molecular Cell Biology*, 4(3):140–152, 2012.

[46] Heinzen, E. L., Arzimanoglou, A., Brashear, A., Clapcote, S. J., Gurrieri, F., Goldstein, D. B., Jóhannesson, S. H., Mikati, M. A., Neville, B., Nicole, S., et al. Distinct neurological disorders with ATP1A3 mutations. *The Lancet Neurology*, 13(5):503–514, 2014.

[47] Heller, M. J. Dna microarray technology: devices, systems, and applications. *Annual Review of Biomedical Engineering*, 4(1):129–153, 2002.

[48] Hirschhorn, J. N. and Daly, M. J. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*, 6(2):95, 2005.

[49] Horvath, I., Jia, X., Johansson, P., Wang, C., Moskalenko, R., Steinau, A., Forsgren, L., Wagberg, T., Svensson, J., Zetterberg, H., et al. Pro-inflammatory S100A9 protein as a robust biomarker differentiating early stages of cognitive impairment in Alzheimers disease. *ACS Chemical Neuroscience*, 7(1):34–39, 2015.

[50] Hossain, S. M. M., Ray, S., and Mukhopadhyay, A. Preservation affinity in consensus modules among stages of HIV-1 progression. *BMC Bioinformatics*, 18(1):181, 2017.

[51] Hosur, R., Xu, J., Bienkowska, J., and Berger, B. iWRAP: an interface threading approach with application to prediction of cancer-related protein–protein interactions. *Journal of Molecular Biology*, 405(5):1295–1310, 2011.

[52] Huang, R., Wallqvist, A., and Covell, D. G. Comprehensive analysis of pathway or functionally related gene expression in the National Cancer Institute's anticancer screen. *Genomics*, 87(3):315–328, 2006.

[53] Huh, W.-K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S., and O'shea, E. K. Global analysis of protein localization in budding yeast. *Nature*, 425(6959):686–691, 2003.

[54] Hung, F.-H. and Chiu, H.-W. Protein-protein interaction prediction based on association rules of protein functional regions. In *Innovative Computing, Information and Control, 2007. ICICIC'07. Second International Conference on*, pages 359–359. IEEE, 2007.

[55] Inoue, H., Takahashi, R., Kondo, T., Iwata, N., and Asai, M. Method for screening therapeutic and/or prophylactic agents for alzheimer's disease, Feb. 20 2013. US Patent App. 14/385,990.

[56] Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences*, 98(8):4569–4574, 2001.

[57] Iwamoto, T., Okumura, S., Iwatsubo, K., Kawabe, J.-I., Ohtsu, K., Sakai, I., Hashimoto, Y., Izumitani, A., Sango, K., Ajiki, K., et al. Motor dysfunction in type 5 adenylyl cyclase-null mice. *Journal of Biological Chemistry*, 278 (19):16936–16940, 2003.

[58] Jang, G.-B., Kim, J.-Y., Cho, S.-D., Park, K.-S., Jung, J.-Y., Lee, H.-Y., Hong, I.-S., and Nam, J.-S. Blockade of Wnt/β-catenin signaling suppresses breast cancer metastasis by inhibiting CSC-like phenotype. *Scientific Reports*, 5, 2015.

[59] Jimenez-Sanchez, G., Childs, B., and Valle, D. Human disease genes. *Nature*, 409(6822):853, 2001.

[60] Kakati, T., Kashyap, H., and Bhattacharyya, D. K. THD-module extractor: an application for CEN module extraction and interesting gene identification for Alzheimers disease. *Scientific Reports*, 6:38046, 2016.

[61] Kanehisa, M. and Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1):27–30, 2000.

[62] Keshava Prasad, T., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., et al. Human protein reference database2009 update. *Nucleic acids research*, 37(suppl_1):D767–D772, 2008.

[63] Keshava Prasad, T., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., et al. Human protein reference database2009 update. *Nucleic Acids Research*, 37 (suppl_1):D767–D772, 2009.

[64] Kikugawa, S., Nishikata, K., Murakami, K., Sato, Y., Suzuki, M., Altaf-Ul-Amin, M., Kanaya, S., and Imanishi, T. PCDq: human protein complex database with quality index which summarizes different levels of evidences of protein complexes predicted from H-Invitational protein-protein interactions integrative dataset. *BMC Systems Biology*, 6(2):S7, 2012.

[65] Kim, S. J., Sung, J. Y., Um, J. W., Hattori, N., Mizuno, Y., Tanaka, K., Paik, S. R., Kim, J., and Chung, K. C. Parkin cleaves intracellular α-synuclein inclusions via the activation of calpain. *Journal of Biological Chemistry*, 278 (43):41890–41899, 2003.

[66] King, A. D., Pržulj, N., and Jurisica, I. Protein complex prediction via cost-based clustering. *Bioinformatics*, 20(17):3013–3020, 2004.

[67] Kitsak, M., Sharma, A., Menche, J., Guney, E., Ghiassian, S. D., Loscalzo, J., and Barabási, A.-L. Tissue specificity of human disease module. *Scientific Reports*, 6, 2016.

[68] Kouhsar, M., Zare-Mirakabad, F., and Jamali, Y. WCOACH: Protein complex prediction in weighted PPI networks. *Genes & Genetic Systems*, 90(5): 317–324, 2015.

[69] Kreek, M. J., Nielsen, D. A., Butelman, E. R., and LaForge, K. S. Genetic influences on impulsivity, risk taking, stress responsivity and vulnerability to drug abuse and addiction. *Nature Neuroscience*, 8(11):1450, 2005.

[70] Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A. P., et al. Global landscape of protein complexes in the yeast saccharomyces cerevisiae. *Nature*, 440(7084):637–643, 2006.

[71] Krumsiek, J., Friedel, C. C., and Zimmer, R. ProCopeprotein complex prediction and evaluation. *Bioinformatics*, 24(18):2115–2116, 2008.

[72] Kumar, A., Agarwal, S., Heyman, J. A., Matson, S., Heidtman, M., Piccirillo, S., Umansky, L., Drawid, A., Jansen, R., Liu, Y., et al. Subcellular localization of the yeast proteome. *Genes & Development*, 16(6):707–719, 2002.

[73] Kumari, S., Nie, J., Chen, H.-S., Ma, H., Stewart, R., Li, X., Lu, M.-Z., Taylor, W. M., and Wei, H. Evaluation of gene association methods for coexpression network construction and biological knowledge discovery. *PLOS ONE*, 7(11):e50411, 2012.

[74] La Creis, R. K., Rogers, E. N., Yeyeodu, S. T., Jones, D. Z., and Kimbro, K. S. Contribution of toll-like receptor signaling pathways to breast tumorigenesis and treatment. *Breast Cancer: Targets and Therapy*, 5:43, 2013.

[75] Lage, K., Karlberg, E. O., Størling, Z. M., Olason, P. I., Pedersen, A. G., Rigina, O., Hinsby, A. M., Tümer, Z., Pociot, F., Tommerup, N., et al. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nature Biotechnology*, 25(3):309–316, 2007.

[76] Latchman, D. S. Transcription factors: an overview. *The International Journal of Biochemistry & Cell Biology*, 29(12):1305–1312, 1997.

[77] Le, D.-H. A novel method for identifying disease associated protein complexes based on functional similarity protein complex networks. *Algorithms for Molecular Biology*, 10(1):14, 2015.

[78] Lee, C.-W., Chang, K.-P., Chen, Y.-Y., Liang, Y., Hsueh, C., Yu, J.-S., Chang, Y.-S., and Yu, C.-J. Overexpressed tryptophanyl-tRNA synthetase, an angiostatic protein, enhances oral cancer cell invasiveness. *Oncotarget*, 6 (26):21979, 2015.

[79] Li, M., Wang, J., and Chen, J. A fast agglomerate algorithm for mining functional modules in protein interaction networks. In *BioMedical Engineering*

and Informatics, 2008. BMEI 2008. International Conference on, volume 1, pages 3–7. IEEE, 2008.

[80] Li, M., Wu, X., Wang, J., and Pan, Y. Towards the identification of protein complexes and functional modules by integrating PPI network and gene expression data. *BMC Bioinformatics*, 13(1):109, 2012.

[81] Li, X.-L., Foo, C.-S., and Ng, S.-K. Discovering protein complexes in dense reliable neighborhoods of protein interaction networks. In *Comput Syst Bioinformatics Conf*, volume 6, pages 157–168. Citeseer, 2007.

[82] Li, Y., Liang, M., and Zhang, Z. Regression analysis of combined gene expression regulation in acute myeloid leukemia. *PLoS Comput Biol*, 10(10): e1003908, 2014.

[83] Lin, L., Göke, J., Cukuroglu, E., Dranias, M. R., VanDongen, A. M., and Stanton, L. W. Molecular features underlying neurodegeneration identified through in vitro modeling of genetically diverse Parkinsons disease patients. *Cell Reports*, 15(11):2411–2426, 2016.

[84] Liu, S. C., Bassi, D. E., Zhang, S. Y., Holoran, D., Conti, C. J., and Klein-Szanto, A. J. Overexpression of cyclin D2 is associated with increased in vivo invasiveness of human squamous carcinoma cells. *Molecular Carcinogenesis*, 34(3):131–139, 2002.

[85] Liu, Z., Yan, S. F., Walker, J. R., Zwingman, T. A., Jiang, T., Li, J., and Zhou, Y. Study of gene function based on spatial co-expression in a high-resolution mouse brain atlas. *BMC Systems Biology*, 1(1):19, 2007.

[86] Lleres, D., Swift, S., and Lamond, A. I. Detecting protein-protein interactions in vivo with FRET using multiphoton fluorescence lifetime imaging microscopy (FLIM). *Current Protocols in Cytometry*, pages 12–10, 2007.

[87] López-Bigas, N. and Ouzounis, C. A. Genome-wide identification of genes likely to be involved in human genetic disease. *Nucleic Acids Research*, 32 (10):3108–3114, 2004.

[88] Lu, L., Lu, H., and Skolnick, J. MULTIPROSPECTOR: an algorithm for the prediction of protein–protein interactions by multimeric threading. *Proteins: Structure, Function, and Bioinformatics*, 49(3):350–364, 2002.

[89] Maere, S., Heymans, K., and Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, 21(16):3448–3449, 2005.

[90] Mahanta, P., Ahmed, H. A., Bhattacharyya, D. K., and Kalita, J. K. An effective method for network module extraction from microarray data. *BMC bioinformatics*, 13(13):S4, 2012.

[91] MathWorks, I. *MATLAB: the language of technical computing. Desktop tools and development environment, version 7*, volume 9. MathWorks, 2005.

[92] Maulik, M., Thinakaran, G., and Kar, S. Alterations in gene expression in mutant amyloid precursor protein transgenic mice lacking niemann-pick type c1 protein. *PloS one*, 8(1):e54605, 2013.

[93] Mazandu, G. K. and Mulder, N. J. DaGO-Fun: tool for Gene Ontology-based functional analysis using term information content measures. *BMC Bioinformatics*, 14(1):284, 2013.

[94] Medina, I. R. and Lubovac-Pilav, Z. Gene Co-Expression Network Analysis for Identifying Modules and Functionally Enriched Pathways in Type 1 Diabetes. *PLOS ONE*, 11(6):e0156006, 2016.

[95] Mewes, H.-W., Frishman, D., Güldener, U., Mannhaupt, G., Mayer, K., Mokrejs, M., Morgenstern, B., Münsterkötter, M., Rudd, S., and Weil, B. MIPS: a database for genomes and protein sequences. *Nucleic Acids Research*, 30(1):31–34, 2002.

[96] Mewes, H.-W., Amid, C., Arnold, R., Frishman, D., Güldener, U., Mannhaupt, G., Münsterkötter, M., Pagel, P., Strack, N., Stümpflen, V., et al. Mips: analysis and annotation of proteins from whole genomes. *Nucleic Acids Research*, 32(suppl_1):D41–D44, 2004.

[97] Mi, H., Poudel, S., Muruganujan, A., Casagrande, J. T., and Thomas, P. D. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Research*, 44(D1):D336–D342, 2015.

[98] Michnick, S. W., Ear, P. H., Landry, C., Malleshaiah, M. K., and Messier, V. Protein-fragment complementation assays for large-scale analysis, functional dissection and dynamic studies of protein–protein interactions in living cells. *Signal Transduction Protocols*, pages 395–425, 2011.

[99] Mukhopadhyay, A., Ray, S., and De, M. Detecting protein complexes in a ppi network: a gene ontology based multi-objective evolutionary approach. *Molecular BioSystems*, 8(11):3036–3048, 2012.

[100] Muller, S., Filippakopoulos, P., and Knapp, S. Bromodomains as therapeutic targets. *Expert reviews in molecular medicine*, 13:e29, 2011.

[101] Nepusz, T., Yu, H., and Paccanaro, A. Detecting overlapping protein complexes in protein-protein interaction networks. *Nature Methods*, 9(5):471–472, 2012.

[102] Ou-Yang, L., Wu, M., Zhang, X.-F., Dai, D.-Q., Li, X.-L., and Yan, H. A two-layer integration framework for protein complex detection. *BMC Bioinformatics*, 17(1):1, 2016.

[103] Pate, K. T., Stringari, C., Sprowl-Tanio, S., Wang, K., TeSlaa, T., Hoverter, N. P., McQuade, M. M., Garner, C., Digman, M. A., Teitell, M. A., et al. Wnt signaling directs a metabolic program of glycolysis and angiogenesis in colon cancer. *The EMBO Journal*, 33(13):1454–1473, 2014.

[104] Peri, S., Navarro, J. D., Amanchy, R., Kristiansen, T. Z., Jonnalagadda, C. K., Surendranath, V., Niranjan, V., Muthusamy, B., Gandhi, T., Gronborg, M., et al. Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Research*, 13 (10):2363–2371, 2003.

[105] Phizicky, E. M. and Fields, S. Protein-protein interactions: methods for detection and analysis. *Microbiological Reviews*, 59(1):94–123, 1995.

[106] Pirooznia, M., Yang, J. Y., Yang, M. Q., and Deng, Y. A comparative study of different machine learning methods on microarray gene expression data. *BMC Genomics*, 9(1):S13, 2008.

[107] Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D., and Brown, P. O. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nature Genetics*, 23(1):41–46, 1999.

[108] Povey, S., Lovering, R., Bruford, E., Wright, M., Lush, M., and Wain, H. The HUGO gene nomenclature committee (HGNC). *Human Genetics*, 109 (6):678–680, 2001.

[109] Povey, S., Lovering, R., Bruford, E., Wright, M., Lush, M., and Wain, H. The hugo gene nomenclature committee (hgnc). *Human Genetics*, 109(6): 678–680, 2001.

[110] Powell, E., Piwnica-Worms, D., and Piwnica-Worms, H. Contribution of p53 to metastasis. *Cancer Discovery*, 4(4):405–414, 2014.

[111] Prasad, T. K., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., et al. Human protein reference database2009 update. *Nucleic Acids Research*, 37 (suppl 1):D767–D772, 2009.

[112] Pu, S., Wong, J., Turner, B., Cho, E., and Wodak, S. J. Up-to-date catalogues of yeast protein complexes. *Nucleic Acids Research*, 37(3):825–831, 2009.

[113] Puccetti, P., Fallarino, F., Italiano, A., Soubeyran, I., MacGrogan, G., Debled, M., Velasco, V., Bodet, D., Eimer, S., Veldhoen, M., et al. Accumulation of an endogenous tryptophan-derived metabolite in colorectal and breast cancers. *PLOS One*, 10(4):e0122046, 2015.

[114] Qi, Y., Balem, F., Faloutsos, C., Klein-Seetharaman, J., and Bar-Joseph, Z. Protein complex identification by supervised graph local clustering. *Bioinformatics*, 24(13):i250–i268, 2008.

[115] Rakshit, H., Rathi, N., and Roy, D. Construction and analysis of the protein-protein interaction networks based on gene expression profiles of Parkinson's disease. *PLOS ONE*, 9(8):e103047, 2014.

[116] Rappaport, N., Nativ, N., Stelzer, G., Twik, M., Guan-Golan, Y., Stein, T. I., Bahir, I., Belinky, F., Morrey, C. P., Safran, M., et al. MalaCards: an integrated compendium for diseases and their annotation. *Database*, 2013: bat018, 2013.

[117] Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., and Barabási, A.-L. Hierarchical organization of modularity in metabolic networks. *Science*, 297 (5586):1551–1555, 2002.

[118] Ray, S. and Maulik, U. Identifying differentially coexpressed module during HIV disease progression: A multiobjective approach. *Scientific Reports*, 7 (1):86, 2017.

[119] Ray, S., Biswas, S., Mukhopadhyay, A., and Bandyopadhyay, S. Detecting Perturbation in Co-Expression Modules Associated with Different Stages of HIV-1 Progression: A Multi-objective Evolutionary Approach. In *Emerging Applications of Information Technology (EAIT), 2014 Fourth International Conference of*, pages 15–20. IEEE, 2014.

[120] Rebhan, M., Chalifa-Caspi, V., Prilusky, J., and Lancet, D. GeneCards: integrating information about genes, proteins and diseases. *Trends in Genetics*, 13(4):163, 1997.

[121] Ruan, J. and Zhang, W. Identifying network communities with a high resolution. *Physical Review E*, 77(1):016104, 2008.

[122] Salwinski, L., Miller, C. S., Smith, A. J., Pettit, F. K., Bowie, J. U., and Eisenberg, D. The database of interacting proteins: 2004 update. *Nucleic Acids Research*, 32(suppl_1):D449–D451, 2004.

[123] Sato, T., Yamanishi, Y., Kanehisa, M., Horimoto, K., and Toh, H. Improvement of the mirrortree method by extracting evolutionary information. *inSequence and Genome Analysis: Method and Applications*, 2011.

[124] Scriver, C. R. and Waters, P. J. Monogenic traits are not simple: lessons from phenylketonuria. *Trends in Genetics*, 15(7):267–272, 1999.

[125] Shamseldin, H. E., Masuho, I., Alenizi, A., Alyamani, S., Patil, D. N., Ibrahim, N., Martemyanov, K. A., and Alkuraya, F. S. GNB5 mutation causes a novel neuropsychiatric disorder featuring attention deficit hyperactivity disorder, severely impaired language development and normal cognition. *Genome Biology*, 17(1):195, 2016.

[126] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11):2498–2504, 2003.

[127] Sharan, R., Ulitsky, I., and Shamir, R. Network-based prediction of protein function. *Molecular Systems Biology*, 3(1):88, 2007.

[128] Sharma, G. and Martin, J. Matlab®: a language for parallel computing. *International Journal of Parallel Programming*, 37(1):3–36, 2009.

[129] Sharma, P. and Bhattacharyya, D. Dcrs: A multi-objective protein complex finding method. In *Proceedings of the International Conference on Computing and Communication Systems*, pages 801–809. Springer, 2018.

[130] Sharma, P., Ahmed, H. A., Roy, S., and Bhattacharyya, D. K. Detecting protein complexes using connectivity among nodes in a PPI Network. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 4(1):35, 2015.

[131] Sharma, P., Ahmed, H. A., Roy, S., and Bhattacharyya, D. K. Unsupervised methods for finding protein complexes from PPI networks. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 4(1):1–15, 2015.

[132] Sharma, P., Bhattacharyya, D. K., and Kalita, J. K. Centrality analysis in PPI networks. In *Accessibility to Digital World (ICADW), 2016 International Conference on*, pages 135–140. IEEE, 2016.

[133] Sharma, P., Bhattacharyya, D. K., and Kalita, J. Disease biomarker identification from gene network modules for metastasized breast cancer. *Scientific Reports*, 7(1):1072, 2017.

[134] Sharma, P., Bhattacharyya, D., and Kalita, J. Detecting protein complexes based on a combination of topological and biological properties in protein-protein interaction network. *Journal of Genetic Engineering and Biotechnology*, 16(1):217–226, 2018.

[135] Shi, L., Lei, X., and Zhang, A. Protein complex detection with semi-supervised learning in protein interaction networks. *Proteome Science*, 9 (1):S5, 2011.

[136] Smith, G. P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, 228:1315–1318, 1985.

[137] Srihari, S. and Leong, H. W. Employing functional interactions for characterisation and detection of sparse complexes from yeast ppi networks. *International Journal of Bioinformatics Research and Applications*, 8(3):286–304, 2012.

[138] Streit, S., Michalski, C. W., Erkan, M., Kleeff, J., and Friess, H. Northern blot analysis for detection and quantification of RNA in pancreatic cancer cells and tissues. *Nature Protocols*, 4(1):37–43, 2009.

[139] Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Muller, J., Bork, P., et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Research*, 39(suppl 1):D561–D568, 2011.

[140] Szymanski, D. B. Breaking the wave complex: the point of arabidopsis trichomes. *Current opinion in plant biology*, 8(1):103–112, 2005.

[141] Tanaka, T., Bai, Z., Srinoulprasert, Y., Yang, B., Hayasaka, H., and Miyasaka, M. Chemokines in tumor progression and metastasis. *Cancer Science*, 96(6):317–322, 2005.

[142] Tassorelli, C., Armentero, M.-T., Greco, R., Fancellu, R., Sandrini, G., Nappi, G., and Blandini, F. Behavioral responses and Fos activation following painful stimuli in a rodent model of Parkinson's disease. *Brain Research*, 1176:53–61, 2007.

[143] Team, R. C. R: A language and environment for statistical computing. *R Foundation for Statistical Computing. Vienna, Austria*, 2017.

[144] Tong, A. H. Y., Lesage, G., Bader, G. D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G. F., Brost, R. L., Chang, M., et al. Global mapping of the yeast genetic interaction network. *Science*, 303(5659):808–813, 2004.

[145] Tong, W., Cao, X., Harris, S., Sun, H., Fang, H., Fuscoe, J., Harris, A., Hong, H., Xie, Q., Perkins, R., et al. Arraytrack–supporting toxicogenomic research at the us food and drug administration national center for toxicological research. *Environmental Health Perspectives*, 111(15):1819, 2003.

[146] van Dam, S., Võsa, U., van der Graaf, A., Franke, L., and de Magalhães, J. P. Gene co-expression analysis for functional classification and gene–disease predictions. *Briefings in Bioinformatics*, page bbw139, 2017.

[147] Van't Veer, L. J., Dai, H., Van De Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., Van Der Kooy, K., Marton, M. J., Witteveen, A. T., et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415(6871):530–536, 2002.

[148] Vanunu, O., Magger, O., Ruppin, E., Shlomi, T., and Sharan, R. Associating genes and protein complexes with disease via network propagation. *PLOS Computational Biology*, 6(1):e1000641, 2010.

[149] Voon, V. and Fox, S. H. Medication-related impulse control and repetitive behaviors in Parkinson disease. *Archives of Neurology*, 64(8):1089–1096, 2007.

[150] Wallén-Mackenzie, Å., Wootz, H., and Englund, H. Genetic inactivation of the vesicular glutamate transporter 2 (VGLUT2) in the mouse: what have we learnt about functional glutamatergic neurotransmission? *Upsala Journal of Medical Sciences*, 115(1):11–20, 2010.

[151] Wang, J. Z., Du, Z., Payattakool, R., Philip, S. Y., and Chen, C.-F. A new method to measure the semantic similarity of GO terms. *Bioinformatics*, 23 (10):1274–1281, 2007.

[152] Wang, K., Li, M., and Hakonarson, H. Analysing biological pathways in genome-wide association studies. *Nature Reviews Genetics*, 11(12):843–854, 2010.

[153] Wang, X., Wang, Z., and Ye, J. HKC: An algorithm to predict protein complexes in protein-protein interaction networks. *BioMed Research International*, 2011, 2011.

[154] Wang, X., Qian, H., and Zhang, S. Discovery of significant pathways in breast cancer metastasis via module extraction and comparison. *IET Systems Biology*, 8(2):47–55, 2014.

[155] Wang, Y., Klijn, J. G., Zhang, Y., Sieuwerts, A. M., Look, M. P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M. E., Yu, J., et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *The Lancet*, 365(9460):671–679, 2005.

[156] Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C. T., et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research*, 38(suppl 2): W214–W220, 2010.

[157] Winkel, B. S. Metabolic channeling in plants. *Annu. Rev. Plant Biol.*, 55: 85–107, 2004.

[158] Wong, D. J. and Chang, H. Y. Learning more from microarrays: insights from modules and networks. *Journal of General Internal Medicine*, 20(5): 175–182, 2005.

[159] Wu, G., Feng, X., and Stein, L. A human functional protein interaction network and its application to cancer data analysis. *Genome Biology*, 11(5): 1, 2010.

[160] Wu, M., Li, X., Kwoh, C.-K., and Ng, S.-K. A core-attachment based method to detect protein complexes in PPI networks. *BMC Bioinformatics*, 10(1): 169, 2009.

[161] Xu, B., Wang, Y., Wang, Z., Zhou, J., Zhou, S., and Guan, J. An effective approach to detecting both small and large complexes from protein-protein interaction networks. *BMC bioinformatics*, 18(12):419, 2017.

[162] Yamada, M., Kabir, M. S., and Tsunedomi, R. Divergent promoter organization may be a preferred structure for gene control in escherichia coli. *Journal of molecular microbiology and biotechnology*, 6(3-4):206–210, 2003.

[163] Yang, Y., Han, L., Yuan, Y., Li, J., Hei, N., and Liang, H. Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. *Nature Communications*, 5:3231, 2014.

[164] Yi, C. and Deng, X. W. Cop1–from plant photomorphogenesis to mammalian tumorigenesis. *Trends in cell biology*, 15(11):618–625, 2005.

[165] Yin, L., Huang, C.-H., and Ni, J. Clustering of gene expression data: performance and similarity analysis. *BMC Bioinformatics*, 7(4):S19, 2006.

[166] Yip, A. M. and Horvath, S. Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinformatics*, 8(1):22, 2007.

[167] Yu, F. Y., Yang, Z. H., Tang, N., Lin, H. F., Wang, J., and Yang, Z. W. Predicting protein complex in protein interaction network-a supervised learning based method. *BMC Systems Biology*, 8(3):S4, 2014.

[168] Yu, G., Li, F., Qin, Y., Bo, X., Wu, Y., and Wang, S. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics*, 26(7):976–978, 2010.

[169] Zaki, N., Efimov, D., and Berengueres, J. Protein complex detection using interaction reliability assessment and weighted clustering coefficient. *BMC bioinformatics*, 14(1):163, 2013.

[170] Zhang, Q., Qin, J., Zhong, L., Gong, L., Zhang, B., Zhang, Y., and Gao, W.-Q. CCL5-mediated Th2 immune polarization promotes metastasis in luminal breast cancer. *Cancer Research*, 75(20):4312–4321, 2015.

[171] Zhang, X.-F., Dai, D.-Q., Ou-Yang, L., and Yan, H. Detecting overlapping protein complexes based on a generative model with functional and topological properties. *BMC Bioinformatics*, 15(1):186, 2014.

[172] Zhao, L. and Zaki, M. J. Tricluster: an effective algorithm for mining coherent clusters in 3d microarray data. In *Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data*, pages 694–705. ACM, 2005.

[173] Zhou, H., Liu, J., Li, J., and Duan, W. A density-based approach for detecting complexes in weighted ppi networks by semantic similarity. *PloS one*, 12(7): e0180570, 2017.

[174] Zotenko, E., Mestre, J., O'Leary, D. P., and Przytycka, T. M. Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality. *PLOS Computational Biology*, 4(8):e1000140, 2008.

Publications based on the Thesis Work

1. **Sharma P.** Ahmed, H. A, Roy, S. and Bhattacharyya, D.K., "Detecting protein complexes using connectivity among nodes in a PPI Network", *NetMAHIB, Springer* vol.4, no.1, pp 1-15, 2015.

2. **Sharma, P.**, and Bhattacharyya, D.K., " DCRS- A multi-objective protein complex finding method", accepted for publication in *International Conference on Computing and Communication Systems, I3CS-2016*

3. **Sharma P.** Bhattacharyya, D.K. and Kalita, J.K, "Centrality Analysis in PPI networks", *International Conference on Accessibility to Digital World (ICADW)*, 2016

4. **Sharma P.** Bhattacharyya, D.K. and Kalita, J.K, "Detecting protein complexes based on a combination of topological and biological properties in protein-protein interaction network", *Journal of Genetic Engineering and Biotechnology, Elsevier*, , 2017 https://doi.org/10.1016/j.jgeb.2017.11.005.

5. **Sharma P.** Bhattacharyya, D.K. and Kalita, J.K, "Protein complex finding and ranking: An application to Alzheimers disease", *Jounal of BioSciences, Springer*, 2016 vol.42, pp-383-396, 2017.

6. **Sharma P.** Bhattacharyya, D.K. and Kalita, J.K, "Disease biomarker identification from gene network modules for metastasized breast cancer", *Scientific Reports, Nature*, doi:10.1038/s41598-017-00996-x (2017).

7. **Sharma P.** Bhattacharyya, D.K. and Kalita, J.K, "Detecting Network Modules to support Biomarker Identification for Parkinson's Disease",[Under Review]