International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER
SSRN

IJIEMR Transactions, online available on 27th Nov 2020. Link :http://www.ijiemr.org/downloads.php?vol=Volume-09&issue=ISSUE-12

DOI: 10.48047/IJIEMR/V09/I12/84

Title: **A SURVEY ON DIFFERENT METHODS USED FOR ANALYSIS OF PROTEIN-PROTEIN INTERACTION NETWORK**

Volume 09, Issue 12, Pages: 446-452

Paper Authors

B Madhav Rao, M Sampath Kumar

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

A SURVEY ON DIFFERENT METHODS USED FOR ANALYSIS OF PROTEIN-PROTEIN INTERACTION NETWORK

[1*] B Madhav Rao, [2] M Sampath Kumar

[1*] Research Scholar, Andhra University, Visakhapatnam, India
madhavraob@gmail.com

[2] Professor, Department of CS&SE, Andhra University College of Engineering(A), Visakhapatnam
[2]*sampathauce@gmail.com*

Abstract: Bio information system is one of the prominent fields for analyzing of biological process. The main objective of Bioinformatics is to identify the disease and analysis the cause for disease. Protein- Protein Interactions (PPI) is used to analyze the structure of protein sequence and visualization in 3D structure. Many methodologies have been used to analysis the cancer causing protein detection using PPI network. Protein-Protein Interaction networks are used for the drug discovery for a particular disease in humans using protein interactions in Human Interaction Networks. There are many advantages and disadvantages while analyzing the different methods, so different analysis and results of large scale data is gathered. It is used for feature directions for the purpose of data modeling and analyzing to be implemented by using different machine learning and deep learning techniques and 3d visualization.Here different analysis methods have been surveyed for the future directions.

Keywords: Bioinformatics, Protein-Protein Interaction, Machine Learning, Deep Learning,3D-Visualization.

1. INTRODUCTION

Biomedical applications are one of the major areas in the research field of information technology to identify biological information for a quality analysis. Integrated genetic and biological information which can be used for the gene based discovery of the dieses by using which a information related to the drugs can be known. It is an area which is a relationship between the computer science and biological sciences. Different biological terms and information techniques are applied to understand the association of molecules on a very large data so this is also termed as management information system for biological sciences. There are many tools for molecular biology which integrates computer science and mathematics to preprocess the biological data for the large dieses information database. These tools perform the complex biological sequences from the large databases at a very fast rate of preprocessing on complex data. The main goals of research in bioinformatics are to maintain the data in such a way which can allow to access the information in a simplest format as per the requirement and retrieve the new information whenever it is produced. So, it is necessary to develop a tool that may be useful for analyzing the data and display the outcome in the human readable format. In the field of Bioinformatics there has been a serious consideration and attentions for the research by using PPI networks. PPI networks can be defined as a result of biochemical or electrostatic forces it forms a connection between two or more proteins. So, the main aim of this paper is to give the different methodologieswhich are used in the areas of colorectal cancer detection using purification of gene expressions [1]. In this paper describes about the various tools used for colorectal cancer PPI network analysis. Most Cellular functionalities are carried out by two or more proteins rather than a single one. Now a days so many Protein-Protein Interaction datasets are publicly available for analysis. Modern computational and mathematical methods required to analysis complex networks formed by combination of two or more proteins [2]. This reduces the cost and time for analyzing the protein sequence with more accuracy than the conventional methods. A valuable framework is required for better interaction and functional organization of the proteome. PPI's are used for performing the biological process of multiple proteins interaction the main goal is unravelling Protein–Protein Interactions in proteomics, which will decode the molecular functionality underlying the biological proteins to understand the human diseases on root level[4]. For detecting protein functionality, it is necessary to analyze the PPI network. By using the analysis target drug is prepared. In PPI network mainly consider about unbound proteins in large number of cells [3].

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

2. IMPORTANCE OF PROTEIN - PROTEIN INTERACTION

Protein–Protein Interaction networks are progressively filling in as devices to decode the atomic premise of illnesses. Moreover, the sequencing of the genome and advances in proteomics prompts the recognizable proof of proteins of obscure capacities. Association organizations may give pieces of information on the elements of these newfound proteins or on new elements of effectively distinguished proteins.

The promising uses of PPI organizations to sickness datasets are focused on four significant regions: (i) the distinguishing proof of qualities and proteins related with illnesses, (ii) the investigation of organization properties and their connection to infection states, (iii) the ID of sickness related subnetworks, and (iv) network-based illness order [10]. Fig. 1 gives a model for the schematic portrayal of understanding illness PPI network relationship utilizing frameworks biomedicine approach.

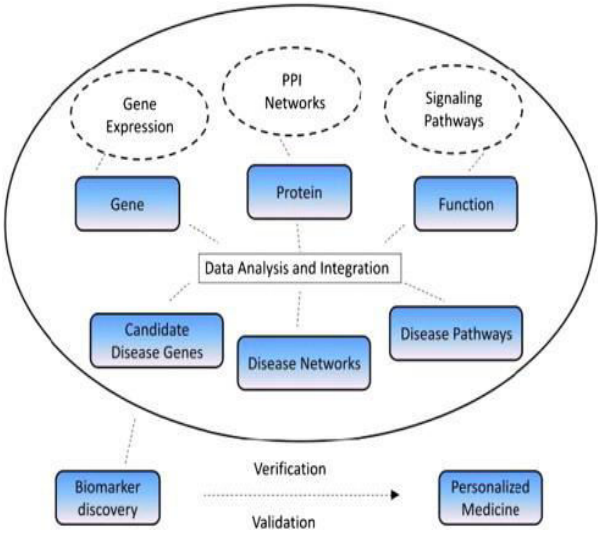

Fig. 1A biomedicine-based approach to understanding the interaction betwe enthe PPI network and disease (Source: Tuba Sevimoglu et al.,2014).

Worldwide comprehension of organizations will permit scientists to inspect the infection pathways and distinguish techniques to control them. The reconciliation of utilitarian genomic and proteomic information to acquire dynamic organization examination will additionally improve the achievement of clinical exploration.

3. PROTEIN-PROTEIN INTERACTION METHODS FOR ANALYZING PROTEINSEQUENCES

Generally, a mix of strategies is critical to validate, characterize and ensure protein interactions. Beforehand unknown proteins could also be found by their association with a number of proteins which are known. Protein interplay evaluation can also uncover unique, unexpected useful roles for well-known proteins. The invention or verification of an interplay is step one on the street to understanding where, how and beneath what circumstances these proteins work together in vivo and the useful implications of those interactions.

3.1 EXPERIMENTAL METHODS

3.1.1 YEAST-TWO-HYBRID SCREENING

This method allows to find interactions between two proteins. In this approach yeast cells are partitioned into two parts, namely prey and bait, prey and bait not interact with each other if no transcription of reporter protein comes. The relation between two proteins is present or not defined by using resultant reporter protein expression [5].In this method there a large scale of false positives and false negatives. Protein folding and expressions are not same in Yeast and humans so that the physical interactions may vary each other.

3.1.2 AFFINITY PURIFICATION OR MASS SPECTROMETRY

Affinity purification or mass spectrometry used to find the constant interactions between proteins and detects the functional relationship between the protein sequences. This technique starts by purification of targeted protein, which is taken. Generally, in this method targeted protein expressed as cell.By using this method PPI can be analyzed quality wise as well as quantity wise [5].

3.1.3GENE EXPRESSION-BASED TECHNIQUES

An active research area is gene expression profiling-based colon cancer detection. A gene will typically undergo three forms of alterations, i.e., over expression, suppression, and mutation of the gene. These improvements have been used for colon cancer identification, and important research studies have been devoted to this field. Genes are commonly studied using various microarray variants, such as microarrays of oligonucleotides and cDNA [28].

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

3.1.4 OLIGONUCLEOTIDE MICROARRAYS

By synthesizing a specific oligonucleotide into a solid surface based on an already established spatial orientation, oligonucleotide microarrays are formed. Using nonfocal laser that analyzes various probes and generates tiff images, oligonucleotide slides are scanned. Images are then processed in order to achieve gene expression levels. For classification, gene expressions are then used.

Alon et al. Examined on an informational collection of 6,500 quality articulations from 22 ordinary and 40 threatening colon tissues in their test, Alon et al. utilized a bunching calculation. An assortment of 2,000 qualities with the most elevated least force across tests has been identified [6]. Contrasted with those in the information assortment, these qualities should be the most oppressive.

Grade et al. chipped away at quality information of 73 threatening and 30 ordinary patients, and discovered 17 discriminative qualities among the information [7]. Additionally, Yajima et al. examined quality articulation profiles of 43 patients (23 reparable malignancy patients—stage A-C, 14 early disease patients—stage A-B, 5 right-sided malignancy patients—stage D) in their exploration study [8]. Quality articulation profiles are acquired from the dung and fringe blood. Three (PAP, REG1A, and DPEP1) and six (SEPP1, RPL27A, ATP1B1, EEF1A1, SFN, and RPS11) most unmistakable qualities are distinguished, separately, in the examples of fringe blood and dung. This arrangement of nine qualities has demonstrated to have the option to precisely distinguish 78 percent of stage A-C, 71 percent of stage A-B, and 80% of stage D patients. Similarly, Kim et al. chipped away at an informational collection of five serrated adenomas and five ordinary colon mucosa tests, and recognized 124 perceiving qualities equipped for recognizing the examples in a viable way [9]. Afterward, Venkatesh et al. proposed another strategy for colon disease identification. Kent Ridge colon disease informational index has been utilized that contains 2,000 quality articulations with most elevated insignificant force across 62 tissues [27].

3.2 COMPUTATIONAL METHODS

3.2.1 TEXTURE ANALYSIS-BASED TECHNIQUES

Surface is a blend of rehashed designs with customary/unpredictable recurrence [11]. There is a critical variety in the surface of typical and threatening colon tissues.

Esgiar et al. proposed a promising strategy for colon malignant growth discovery by utilizing textural highlights. In their work, unique colon biopsy pictures of size 512 are additionally separated into four subimages of size 256, and the subimages having little tissue content are prohibited. Dark level cooccurrence lattice (GLCM) is then determined for each subimage. Standardized GLCM is utilized to decide textural highlights of precise second, contrast, relationship, converse distinction second, difference, and entropy [12]. The detailed characterization exactness is 90.2 percent for a mix of relationship and entropy by utilizing straight separate examination (LDA) classifier.

Esgiar et al. further expanded their previous work, and used numerical and surface features. Numerical features include features of shape and course. Surface features incorporate energy, lethargy and homogeneity, and are resolved from GLCM of the image. The declared request precision for numerical and surface features is 80 and 90%, independently [13].

3.2.2 OBJECT ORIENTED TEXTURE ANALYSIS-BASED TECHNIQUES

These methodologies abuse establishment data about size and spatial scattering of colon tissue sections for division and request of colon biopsy pictures. These systems have been furthermore isolated into division and portrayal strategies.

Object Oriented Texture Analysis-Based Segmentation Techniques Initially, OO surface examination was planned for division of colon biopsy pictures. In this association, object-oriented division (OOSEG) is the main OO surface examination based division method that includes three very much characterized stages, specifically, object definition, surface definition, and division [14]. In item definition stage, k-means is applied to separate picture pixels into three bunches relying on shading forces of tissue segments, for example, purple-hued cores, white-hued lumen and epithelial cells, and pink-hued stroma [19].

OOSEG however creates sensible outcomes has a couple of constraints. For example, it needs manual change of boundaries for each test picture [14]. Subsequently, Tosun et al. proposed another method with an intend to recognize a bunch of boundary esteems material to all picture examples. In this method, object finding measure is like OOSEG, however estimations in later stages are performed regarding objects instead of pixels. In element extraction stage, a bunch of twelve highlights as utilized in is determined for each item. Next, Voronoi

outline is developed on the centroids of the multitude of objects to decide beginning seeds. Any two nearby articles are assembled if Euclidean distance between them is more modest than a predefined likeness limit. Afterward, bunches having number of articles bigger than a limit are announced seeds [14] [15].

Demir et al. proposed a significant colon biopsy picture division procedure. In this method, circle fitting cycle is like OOSEG with one exemption that solitary purple and white groups are utilized to find core and lumen objects, individually [14] [16]. Next, an item chart is developed on these articles. Edges are doled out between every lumen item and its N nearest lumen and N nearest core neighbors. For every lumen object L, highlights having data about zones, length of edges among L and its core and lumen neighbors are separated by thinking about neighbors inside a round window around L. These highlights are additionally utilized by the k-implies calculation to isolate lumen objects into "organ" and "nongland" classes [20]. Objects of "organ" class are treated as beginning seeds. Locale developing cycle includes another item diagram that is built on core objects. Beginning from the underlying lumen seeds, more lumen objects are added to the diagram until an edge of the core chart is experienced. Edges of the core diagram are utilized to stop district developing since organs are generally enclosed by the core protests, and experiencing a core object implies that organ limit is reached. Eventually, bogus organ disposal, which has been demonstrated to be successful in and, is applied to eliminate bogus organs [24, 25].

Tosun et al. further improved OOSEG by utilizing charts for evaluating spatial connection between cytological tissue segments [17]. In the initial step of this work, recently proposed techniques for circle fitting and diagram age are utilized [16][20]. In the subsequent advance, diagram edge runs are determined. Chart edge runs depend on dim level run-length grids.Graph-edge run is a path that starts from an initial node, and contains all nodes reachable with a set of edges of the same type. For computation of dim level run-length network (GRLM), a roundabout window is speculated at focus of a hub, and afterward broadness first inquiry is utilized to process way for every specific edge type that exists in the window[21]. In feature extraction phase, four features, namely, short-path emphasis (SPE), long-path emphasis (LPE), edge type nonuniformity (ETN), and path length nonuniformity (PLN) are computed. ETN and PLN help in deciding the impact of edge type and way length appropriation on surface, and possess least qualities when the runs are consistently conveyed over all edge types and way lengths. Division thoroughly depends on items rather than pixels, and contains three stages: seed assurance, area developing, and district combining. A window is fixated on current item, and amassed GRLM of the enclosed article is determined, which is utilized in element figuring of current article. Beginning seeds are controlled by disengaging sets of contiguous articles having middle of the road distance more prominent than a distance limit, and eliminating parts having lesser number of items than a predefined edge. Items are converged to the seeds on the off chance that they are nearby, and euclidean distance between their highlights is more modest than combine limit. At last, Voronoi outline of the items is built to delineate last area limits.

Simsek et al. acquainted cooccurrence highlights with measure spatial connection between objects in a colon biopsy picture [18]. Roundabout items are situated by utilizing circle fitting calculation [14]. A cooccurrence network is determined for each item by putting a round window on the article, and estimating the occasions objects of one kind cooccur with objects of one more sort at a given distance d. 24 cooccurrence highlights are removed from the cooccurrence framework. In this method, division has been acted like a diagram parceling issue. Irregular items are picked in various cycles to produce charts. Division is accomplished by utilizing these charts. At long last, numerous outcomes are joined to acquire last division.

NivesSkunca et al., team explain the Phylogenetic Profiling. It detects the similar patterns in protein families of large number of organisms. This method assumes, that if there is an interaction between proteins then they must co-evolve. In this method, phylogenetic profile is generated for each protein. By using this method, this method can be used to discover the pathways for unknown enzymes [NivesSkunca et al., 2015][22] [23].

4. PRIMARY DATABASES

Protein-Protein Interaction networks form an interaction among proteins that may be single interaction or hundreds of interactions. All this data has been collected from different databases. These are specifically designed for biological data storage. These databases are frequently updated in order to provide accurate and complete interaction data to users. Day by day biological data and databases increases. Protein Databases are classified into three categories, they are prediction, primary and Meta databases.The following

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

table shows various protein interaction databases available for analysis.

Table-1: Different Protein Databases Available for Protein Analysis

DATABASE	INTER-ACTIONS	ORGA-NISMS	LAST UPDATE	VERSION	WEBSITE
BioGRID	16,70,339	68	MAR 2019	3.5.170	https://thebiogrid.org/
STRING-DB	312,30,56,667	5,090	JAN 2019	11.0	https://string-db.org/
DIP	81,923	834	FEB 2017	--	https://dip.doembi.ucla.edu/dip/Main.cgi
IntAct	8,75,852	09	FEB 2019	4.2.12	https://www.ebi.ac.uk/intact
MINT	1,26,712	645	FEB 2018	---	https://mint.bio.uniroma2.it/
Innatedb	3,67,527	5	JAN 2019	5.4	https://www.innatedb.com/
HPRD	41,327	HUMANS	APR 2010	9	http://hprd.org/
EcoCyc	36,151	E. coli	DEC 2018	22.6	https://ecocyc.org/
CORUM	8,75,852	HUMAN, RAT, MOUSE	SEPT 2018	3.0	http://mips.helmholtz-muenchen.de/corum
BINDINGDB	15,58,402	--	NOV 2017	--	http://www.bindingdb.org/bind/index.jsp

5. TOOLS USED FOR ANALYSIS OF PPI NETWORKS

There are many analyzing tools available which are used to analyze the Protein-Protein Interaction Network. Some of the analyzing tools are:

Cytoscape is an open source tool for constructing small as well as complex PPI networks and as it analyzes the network. It can be merged multiple protein networks into single one. It produces 3D structure representation of PPI network. It supports predefined tools as well as user defined tools to analyze the PPI network.

PathBLAST tool is PPI network analyzer and search tool for comparing PPI network among different species to detect the pathway of the proteins.

BiogridPlugin2 is used to importing data from BioGrid Database. Here various filters can be applied to get specific PPI network to import.

iFrag predicts binding between two protein arrangements utilizing scoring matrix for comparative buildup sections existing at the interface area of the proteins. A server has been actualized to encourage the utilization of iFrag [26]. The contribution of iFrag is in two groupings in FASTA position (inquiry proteins). The client can adjust the BLAST to seek conditions by determining the greatest E-esteem limit and inclusion of the arrangements with the inquiry. The arrangement of layout co-operations can be chosen by the technique for test location (for example by barring co-complex techniques).

6. CONCLUSIONS

Whereas accessible strategies are unable to foretell interactions with 100% accuracy, computational strategies will scale down the set of potential interactions to a subset of most definitely interactions. These interactions will function a place to begin for furthermore lab experiments. For analyzing Protein-Protein interactions various methods are available in data mining, soft computing in computational methods and experimental methods like Y2H, co-IP. Various databases available for Protein-Protein Interaction datasets. In these databases some are based on structure of the protein and some are sequence based. Experimental methods gives less accuracy when compared machine learning techniques discussed in this paper. There is a possibility for analyzing large scale protein datasets by using deep learning techniques.This survey paper gives overview of methods available in PPI network analysis.

REFERENCES

[1] Huang D-S, Zhang L, Han K, Deng S, YangK, Zhang H. Prediction of Protein-ProteinInteractions based on Protein- Proteincorrelation using least squares regression.Curr Protein Pept Sci. 2014;15:553–560

[2] Antonio Mora, Katerina Michalickovaand Ian M Donaldson, “A survey ofprotein interaction data and multigenicinherited disorders”, BMCBioinformatics, vol. 14, pp. 1-7, 2013.

[3] Joan Planas-Iglesias,JaumeBonet,JavierGarcia-Garcia,Manuel A. Marin-Lopez, ElisendaFelliuandBaldoOliva,”Understanding Protein–ProteinInteractions Using Local StructuralFeatures”, JMB Article, pp.1-3, 2013.

[4] Hung Xuan Ta and LiisaHolm,“Computational approaches for predictingprotein interaction

networks: The wiringof protein", The BiochemicalSociety,pp.1- 3,2011., S. Mullender

[5] Anna Brückner , Cécile Polge , NicolasLentze, Daniel Auerbach and UweSchlattner," Review Yeast Two-Hybrid, aPowerful Tool for Systems Biology"; Int. J.Mol. Sci. 2009, 10, 2763-2788;doi:10.3390/ijms10062763

[6] U. Alon, N. Barkai, and D.A. Notterman, "Broad Patterns of GeneExpression Revealed by Clustering Analysis of Tumor andNormal Colon Tissues Probed by Oligonucleotide Arrays," Proc.Nat'l Academy of Science USA, vol. 96, pp. 6745-6750, 1999.

[7] M. Grade, P. Ho¨rmann, and S. Becker, "Gene Expression ProfilingReveals a Massive, Aneuploidy-Dependent Transcriptional Deregulationand Distinct Differences between Lymph Node-Negative and Lymph Node-Positive Colon Carcinomas," CancerResearch, vol. 67, pp. 41-56, 2007.

[8] S. Yajima et al., "Expression Profiling of Fecal Colonocytes forRNA-Based Screening of Colorectal Cancer," Int'l J. Oncology,vol. 31, pp. 1029-1037, 2007.

[9] K. Kim, U. Park, and J. Wang, "Gene Profiling of ColonicSerrated Adenomas by Using Oligonucleotide Microarray,"Int'l J. Colorectal Diseases, vol. 23, pp. 569-580, 2008.

[10] Tuba Sevimoglu, Kazim YalcinArga, "The role of protein interaction networks in systems biomedicine", Computational and Structural Biotechnology Journal, Volume 11, Issue 18, Pages 22-27, August 2014.

[11] S. Rathore et al., "Texture Analysis for Liver Segmentation andClassification: A Survey," Proc. Frontiers of Information Technology,pp. 121-126, 2011.

[12] A.N. Esgiar et al., "Microscopic Image Analysis for QuantitativeMeasurement and Feature Identification of Normal and CancerousColonic Mucosa," IEEE Trans. Information Technology inBiomedicine, vol. 2, no. 3, pp. 197-203, Sept. 1998.

[13] A.N. Esgiar et al., "Automated Feature Extraction and Identificationof Colon Carcinoma,"Analytical and Quantitative Cytology andHistology, vol. 20, pp. 297-301, 1998.

[14] A.B. Tosun et al., "Object-Oriented Texture Analysis for theUnsupervised Segmentation of Biopsy Images," J. Pattern Recognition, vol. 42, pp. 1104-1112, 2009.

[15] A.B. Tosun, C. Sokmensuer, and C.G. Demir, "UnsupervisedTissue Image Segmentation through Object-Oriented Texture,"Proc. 20th Int'l Conf. Pattern Recognition, pp. 2516-2519, 2010.

[16] C.G. Demir et al., "Automatic Segmentation of Colon GlandsUsing Object-Graphs," Medical Image Analysis, vol. 14, pp. 1-12,2010.

[17] A.B. Tosun and C.G. Demir, "Graph Run-Length Matrices forHistopathological Image Segmentation," IEEE Trans. MedicalImaging, vol. 30, no. 3, pp. 721-732, Mar. 2011.

[18] A.C. Simsek et al., "Multilevel Segmentation of HistopathologicalImages Using Cooccurrence of Tissue Object," IEEE Trans.Biomedical Eng., vol. 59, no. 6, pp. 1681-1690, June 2012.

[19] K. Fukunaga and L. Hostetler, "The Estimation of the Gradient of aDensity Function, with Applications in Pattern Recognition," IEEETrans. Information Theory, vol. 21, no. 1, pp. 32-40, Jan. 1975.

[20] C.G. Demir, "Mathematical Modeling of the Malignancy of CancerUsing Graph Evolution," Math. Biosciences, vol. 209, no. 2, pp. 514-527, 2007.

[21] M.M. Galloway, "Texture Analysis Using Gray Level RunLengths," Computer Graphics and Image Processing, vol. 4, no. 2, pp. 172-179, 1975.

[22] "Colon Cancer Data Set Kent Ridge," http://datam.i2r.astaredu.sg/datasets/krbd/ColonTumor/ColonTumor.html",2013.

[23] D. Myers, "Colon Cancer Stages: Basics of Each Colon CancerStage,"http://coloncancer.about.com/od/stagesandsurvivalrate1/a/ColonCancerStag.html", Jan. 2013.

[24] H.S. Wu et al., "Segmentation of Intestinal Gland Images withIterative Region Growing," J. Microscopy, vol. 220, no. 3, pp. 190-204, 2005.

[25] S. Naik et al., "Gland Segmentation and Gleason Grading ofProstate Histology by Integrating Low-High-Level and DomainSpecific Information," Proc. Second Workshop Microscopic ImageAnalysis with Applications in Biology, 2007.

[26] Garcia-Garcia J, Valls-Comamala V, Guney E, Andreu D, Munoz F. J, Fernandez-Fuentes N and Oliva B, "iFrag: A Protein-Protein Interface Prediction Server Based on Sequence Fragments", Journal of Molecular Biology, 429(3): 382–389. doi:10.1016/j.jmb.2016.11.034, 2017.

[27] E.T. Venkatesh, P. Thangaraj, and S. Chitra, "An Improved Neural Approach for Malignant and Normal Colon Tissue Classification from

Oligonucleotide Arrays," European J. Scientific Research, vol. 54, pp. 159-164, 2011.

[28] Franc¸ois Bertucci1 et al.,2014) Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters, pp:1377-1396.