Bioinformatics, Proteomics and Immaging Analysis

OMMEGA Publishers
Answering Intelligence
www.ommegaonline.com

Research Article

InterCluster-A Tool to Cluster Protein-Protein Interactions: Datamining of Protein Interactions in Primary Open Angle Glaucoma

Sudhakaran Dhanya, Eswari PJ Pandaranayaka*

Center of Excellence in Bioinformatics, School of Biotechnology,Madurai Kamaraj University, Madurai – 625021, India

Corresponding authors: Pandaranayaka, E.P.J. Center of Excellence in Bioinformatics, School of Biotechnology, Madurai Kamaraj University, Madurai – 625021, India. E-mail: eswari@mkustrbioinfo.com, eswaripj@gmail.com

Abstract

A growing number of diseases seem to be associated with protein aggregation and each disease has several proteins involved in it. To obtain a better understanding of the diseases, the proteins involved in them and their primary interaction partners were collected and clustered. A tool is developed to aid in clustering these proteins and all their primary interactors. The tool is used to cluster proteins involved in Primary Open Angle Glaucoma and their primary interactors. A cluster was selected for analysis based on the availability of experimental analysis in literature. The localization of the proteins in the chosen cluster was collected. On analyzing, four of the proteins in the cluster was found to be associated with heparin binding. Primary open angle glaucoma is known to be associated with loss of retinal vasculature. The tool has helped in finding a cluster of protein interactions with more experimental data. Also it has helped in finding out the 4 proteins associated with the disease that are involved in heparin binding from 10500 proteins. This would not have been possible to do manually. Further studying the role of these four proteins based on heparin binding and loss of vasculature in primary open angle glaucoma would give a better understanding of the disease and the molecular mechanism involved in it.

Received Date: May25, 2015
Accepted Date: July 14, 2015
Published Date: July 16, 2015

Citation: Dhanya, S., et al. InterCluster-A Tool to Cluster Protein-Protein Interactions: Datamining of Protein Interactions in Primary Open Angle Glaucoma. (2015) Bioinfo Proteom Img Anal 1(1): 15- 19.

DOI: 10.15436/2381-0793.15.003

Introduction

Protein aggregation and disease

Proteins are macromolecules responsible for a wide array of functions in every living organism. Right interactions of proteins either with other proteins or other macromolecules or drugs are crucial for the healthy functioning of any organism. When there is modification or disruption in these interactions, proteins tend to aggregate leading to several diseases (http://bicmku.in/ProADD) like Alzheimer, Parkinson, Huntington, Amyotropic Lateral Sclerosis, Spongiform encephalopathy, Drepanocytosis, Type II Diabetes, Primary Open Angle Glaucoma (POAG), Blepharophimosis Ptosis Epicanthus Inversus Syndrome (BPES)[1-4]. Protein aggregation occurs either due to mutations, misfolding and/or intrinsic disorder in proteins[5-7]. Proteins like β 2 microglobulin, transthyretin, prion protein, lysozyme aggregate due to misfolding. But proteins like amylin, alphasynuclein, β amyloid peptide aggregate due to intrinsic disorder[7]. Mutations in proteins like prion protein, hemoglobin, superoxide dismutase, forkhead domain containing proteins and myocilin, play role in protein aggregation[1,4,5,8,9]. Molecular mechanisms involved in the transition of folded biologically-functional protein molecules into aggregates remain poorly understood. In certain cases as in amyloid and prion diseases, protein aggregation follows a pattern, but the exact mechanism is not clear yet[10]. Studying the protein-protein interactions prevailing in these diseases might help in understanding the molecular mechanism. A tool InterCluster (http://bicmku.in/InterCluster) has been developed to cluster the protein-protein interactions in one of the protein aggregation diseases - POAG and analyzed.

Materials and Methods

Proteins involved and primary interactors

The list of proteins involved in POAG was collected from literature and UniProt database (www.uniprot.org). Ten proteins were obtained from UniProt. Literature search was also done to collect details of other proteins involved in POAG. The availability of 3D structure was checked for every protein from PDB (www.rcsb.org). The primary interactors of these proteins were then collected from STRING database (www.string-db.org). The confidence score was set as 0.150 and the maximum number of interactors was limited to 500 by the database. The resultant lists of primary interactors were saved in a text file.

Localization and function

The localization information of the proteins involved in the disease was collected using CoPub Portal (www.copub.org). Function of the proteins was obtained from UniProt and confirmed using literature search.

Tool development

A tool was developed to cluster the proteins involved in the disease and their primary interactors. Python was used as the back end and the web interface was designed using Hypertext Preprocessor (PHP) and Hyper Text Markup Language (HTML). The web interface allows the user to upload the list of primary interactors in a compressed format. The input to the tool is a compressed folder which consists of list of primary interactors, stored in a text file named with the corresponding protein. Thus there are text files, with list of primary interactors for every protein. The tool results in clusters at all levels. The five output files generated are explained below.

Occurrence of primary interactor among the proteins: The occurrence of primary interactor among the proteins is given in the file "Name.out". This includes the list of primary interactors along with the protein with which they interact.

Interaction number of the primary interactors: The number of interactions for each primary interactor is given in the file "Num.out".

Clusters within interaction group: The cluster information at all levels is in the file "Cluster.out". Under each "Interaction number" group, there are various sub-groups and clusters in each of them are shown. X>Y implies that, there are X primary interactors common to Y proteins involved in the disease.

Consolidated list of all clusters: The consolidated list of all clusters is presented in "List.out". The clusters are obtained in the following format:

(IN : SN) I1, I2, I3 --> P1, P2, P3[X>Y]

The numbers IN and SN corresponds to interaction number and sub-group number (Interaction number: Sub-group number). The list of primary interactors (I1,I2,I3) is on the left hand side and the proteins involved in the disease (P1,P2,P3) are on the right hand side. The numbers at the end corresponds to the number of primary interactor and the number of proteins involved in the disease. [Number of primary interactor (X)> Number of proteins involved in the disease (Y)].

Detailed information about the clusters: The detailed information about the clusters is given in "Output.out". This file gives elaborate details about the cluster and how the cluster is formed.

Validation of the tool

To validate the tool, a dataset was created, based on T-cell receptor pathway. The proteins on the membrane were considered equivalent to the proteins involved in the disease. The proteins directly interacting with these membrane proteins were considered as the primary interactors. The files were then given as input to the tool.

Results

Proteins and their primary interactors

POAG is a major cause of blindness, characterized by progressive degeneration of the optic nerve and is usually associated with elevated intraocular pressure (IOP). This results in loss of retinal ganglion cell axons, along with supporting glia and vasculature. Reducing the intraocular pressure prevents progression of the disease in all stages[11,12].

Thirty one proteins are involved in POAG, fifteen of them have three dimensional structure (Table 1). The number of primary interactors varied from tens to five hundreds. The total number of primary interactors for all the proteins together was around 10500.

Table 1: Proteins involved in primary open angle glaucoma.

SL NO	STRING Database name	PROTEIN NAME
1	CELF2	CUGBP Elav-like family member 2
2	CYP1B1*	Cytochrome P450 1B1
3	MYOC	Myocilin
4	NPHP4	Nephrocystin-4
5	NTF4 *	Neurotrophin-4
6	OPTN *	Optineurin
7	RPGR1P1	X-linked retinitis pigmentosa GTPase regulator-interacting protein 1
8	WDR36	WD repeat-containing protein 36
9	TBK1 *	Serine/threonine-protein kinase TBK1
10	ASB10	Ankyrin repeat and SOCS box protein 10
11	OCLM	Oculomedin
12	ELN	Elastin
13	FN1 *	Fibronectin
14	SPARCL1	SPARC-like protein 1 (hevin)
15	LOXL1	Lysyl oxidase homolog 1
16	CAV1	Caveolin-1
17	CAV2	Caveolin-2
18	APOE *	Apolipoprotein E
19	OLFM1	Noelin
20	OLFM2	Noelin-2
21	OLFM3	Noelin-3
22	TTR *	Transthyretin
23	NPPA *	Natriuretic peptides A
24	OPA1	Optic atrophy 1
25	TP53 *	Cellular tumor antigen p53
26	GSTM1 *	Glutathione S-transferase Mu 1
27	GSTT1 *	Glutathione S-transferase theta-1
28	IL1A *	Interleukin-1 alpha
29	IL1B *	Interleukin-1 beta
30	COL15A1 *	Collagen alpha-1 (XV) chain
31	COL18A1 *	Collagen alpha-1 (XVIII) chain

* proteins with pdb entry

InterCluster

InterCluster is a web-based server for clustering proteins and primary interactors. InterCluster is freely accessible at http://bicmku.in/InterCluster (Figure 1). The total number of

proteins involved in POAG and their primary interactors from STRING database is around 10500. InterCluster clusters these proteins and results in 16 groups and 110 clusters. The cluster with the maximum interaction number [7 > 11] is selected as Cluster 1. The [9 > 3] sub-cluster in the [11:9] cluster is considered as Cluster 2 as it has maximum experimental evidence.

Figure 1: Home page of InterCluster.

Validation of the tool

The tool was validated using primary interactions in T-cell receptor pathway. The tool was able to group the primary interactors and the proteins to form clusters as involved in the pathway (Figure 2). For example, 2 membrane proteins in the pathway PD-1 and CTLA4 interact with the protein SHP1. Here, SHP1 is considered as a primary interactor. Hence the cluster obtained is: SHP1 > PD-1, CTLA4

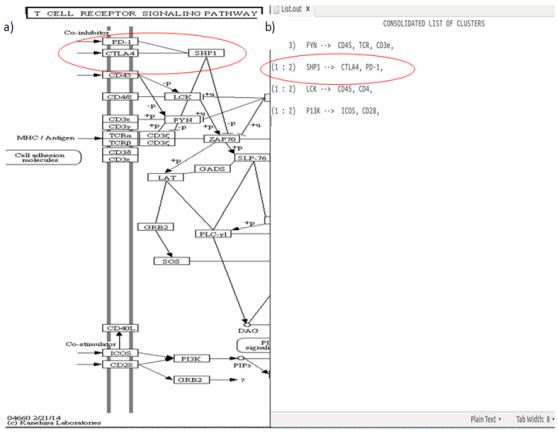

Figure 2: Validation of the tool. a) Part of T-cell receptor pathway used for validation. (KEGG : www.genome.jp/kegg) b) Clusters obtained from InterCluster.

Clusters from InterCluster

Cluster 1: (17:7) POLA1, TNF, ALB, GAPDH, POLD1, TSPO, HRAS --> APOE, CYP1B1, TBK1, GSTM1, ELN, OPA1, MYOC, IL1A, FN1, IL1B, NPPA [7>11]

This cluster has the maximum number of interaction. Eleven proteins involved in POAG have seven primary interactors (Table 2) in common. The functions of the proteins involved in this cluster were quite diverse. From the localization data acquired, some of the POAG proteins and the primary interactors were found to be co-localized. For example, MYOC (Uniprot ID:Q99972) and HRAS (UniProt ID:P01112) are co-localised in mitochondria[13,14]. Studies also show that both MYOC and HRAS play a role in cell proliferation and survival[15,16]. Further analysis was not carried out due to lack of experimental data in literature.

Table 2: Primary interactors in Cluster 1

SL NO	STRING Database name	PROTEIN NAME
1	GAPDH *	Glyceraldehyde-3-phosphate dehydrogenase
2	TNF *	Tumor necrosis factor
3	HRAS *	Intrinsic GTPase activity
4	POLA1 *	DNA polymerase alpha catalytic subunit
5	POLD1	DNA polymerase delta catalytic subunit
6	ALB *	Albumin
7	TSPO	Translocator protein

* proteins with pdb entry

Cluster 2: (11:9) ELANE, STAT3, TIMP1, EDN1, KNG1, MYOC, LDLR, WDFY2, TF --> FN1, APOE, CYP1B1[9>3]

This cluster was chosen for analysis as this had maximum number of experimental analysis. Three proteins involved in POAG has nine primary interactors (Table 3) in common. In this cluster, four proteins, MYOC (Uniprot ID: Q99972), APOE (UniProt ID: P02649), FN1 (UniProt ID: P02751) and CYP1B1 (UniProt ID: Q16678) are proteins involved in POAG. Though MYOC is a protein involved in POAG, in this cluster it occurs only as a primary interactor. It is known that mutations in MYOC and APOE cause POAG[17]. Accumulation of FN1 in the trabecular meshwork is associated to POAG[18]. Mutant CYB1B1 causes MYOC upregulation, which in turn causes POAG pathogenesis[19].

Table 3: Primary interactors in Cluster 2

SL NO	STRING Database name	PROTEIN NAME
1	ELANE*	Neutrophil elastase
2	STAT3	Signal transducer and activator of transcription 3
3	TIMP1*	TIMP metallopeptidase inhibitor 1
4	EDN1*	Endothelin-1
5	KNG1*	Kininogen-1
6	MYOC*	Myocilin
7	LDLR*	Low-density lipoprotein receptor
8	WDFY2	WD repeat and FYVE domain-containing protein 2
9	TF*	Transferrin

* proteins with pdb entry

Clinical studies show that POAG is significantly more prevalent in a group of people with optic cup - retinal venous occlusion (OC-RVO). RVO is the blockage of the small veins that carry blood away from the retina. The mean IOP is significantly higher in OC-RVO than in other types of RVO[20]. POAG is also associated with increase in IOP and loss of vasculature[11]. From this analysis the 4 proteins APOE, MYOC, ELANE (UniProt ID: P08246) and KNG1 (UniProt ID: P01042) in this cluster are found to be involved in heparin binding.[21-24]

Further, based on the localization of the proteins[25-31]

in Cluster 2 (Figure 3) it is most probable that STAT3 (UniProt ID: P40763) interacts with POAG proteins APOE and CYP1B1 in the cytoplasm. Also MYOC might interact with POAG proteins FN1 and CYP1B1 in the endoplasmic reticulum. The localization information of these proteins gathered from literature is based on experimental analysis.

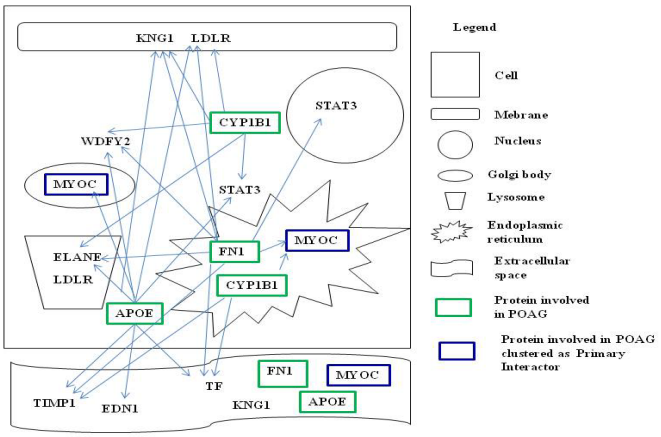

Figure 3: Localisation and interactions of the proteins in cluster 2. The proteins in the cluster 2 are placed in their subcellular locations and the arrows indicate the interactions between them based on literature.

Discussion

POAG is associated with IOP, degeneration of the optic nerve and loss of vasculature[11]. InterCluster, a tool developed to cluster protein-protein interactions, has helped in datamining the proteins involved in POAG and their primary interactions. Among 10500 proteins that were primary interactors of the 31 proteins involved in POAG we could datamine four proteins APOE, MYOC, ELANE and KNG1 in a cluster (Cluster 2) experimentally proved to be involved in heparin binding[21-24]. This would have been impossible without the aid of the tool. These proteins may probably be involved in the pathogenesis of POAG by playing a major role in RVO leading to increase in IOP and loss of vasculature which is one of the reasons for neuronal death in POAG[11]. Understanding the protein-protein interactions involved in POAG and other protein aggregation diseases would also help in elucidating protein-drug interactions as in the case of clozapine induced agranulocytosis[32] which would help in finding appropriate drugs to find cure for these incurable diseases. Pathogenesis on POAG has been a mystery though decades of research has been carried out as only 2-3% of the disease is associated with mutation. With increasing research more proteins are known to be involved in the disease based on mutation analysis. But the real reason has not been yet understood. Our study has opened a new way to look at the pathogenesis of POAG and further elaborate research in the new direction.

Acknowledgement

Department of Biotechnology, New Delhi for facilities at the Center of Excellence in Bioinformatics and fellowship to SD. University Grants Commission, New Delhi is acknowledged for Dr. D. S. Kothari Post-Doctoral fellowship to EPJP.

References

1. Watanabe, M., Dykes-Hoberg, M., Culotta, V. C., et al. Histological Evidence of Protein Aggregation in Mutant SOD1 Transgenic Mice and in Amyotrophic Lateral Sclerosis Neural Tissues. (2001) Neurobiol Dis 8(6): 933-941.

2. Aguzzi, A., O'Connor, T. Protein aggregation diseases: pathogenicity and therapeutic perspectives. (2010) Nat Rev Drug Discov 9(3):237-248.

3. Todeschini, A.L., Dipietromaria, A.,L'Hote, D., et al. Mutational probing of the forkhead domain of the transcription factor FOXL2 provides insights into the pathogenicity of naturally occurring mutations. (2011) Hum Mol Genet 20(17):3376-3385.

4. Pandaranayaka, P.J., Prasanthi, N., Kannabiran, N., et al. Polymorphisms in an intronic region of the myocilin gene associated with primary open-angle glaucoma--a possible role for alternate splicing. (2010) Mol Vis16:2891-2902.

5. Nallathambi, J., Laissue, P., Batista, F., et al. Differential functional effects of novel mutations of the transcription factor FOXL2 in BPES patients. (2008) Hum Mutat 29:123-131.

6. Kanagavalli, J., Krishnadas, S.R., Pandaranayaka, E., et al. Evaluation and understanding of myocilin mutations in Indian primary open angle glaucoma patients. (2003) Mol Vis 9:606-614.

7. Groot, N.S., Pallares, I., Aviles, F.X., et al. Prediction of "hot spots" of aggregation in disease-linked polypeptides. (2005) BMC Struct Biol 5:18.

8. Bunn, H.F., Noguchi, C.T., Hofrichter, J., et al. Molecular and cellular pathogenesis of hemoglobin SC disease. (1982) Proc Natl Acad Sci U S A 79(23):7527-7531.

9. Guo, J., Ning, L., Ren, H., et al. Influence of the pathogenic mutations T188K/R/a on the structural stability and misfolding of human prion protein: Insight from molecular dynamics simulations. (2012) Biochim Biophys Acta 1820(2):116-123.

10. Kumar, S., Udgaonkar, J.B. Mechanisms of amyloid fibril formation by proteins. (2010) Curr Sci 98(5): 639-656.

11. Kwon, Y.H., Fingert, J.H., Kuehn, M.H., et al. Primary Open-Angle Glaucoma. (2009) N Engl J Med 360(11):1113-1124.

12. Weinreb, R.N., Khaw, P.T.Primary open-angle glaucoma. (2004) Lancet 363(9422):1711-1720.

13. Wentz-Hunter, K., Ueda, J., Shimizu, N., et al. Myocilin is associated with mitochondria in human trabecular meshwork cells. (2002) J Cell Physiol 190(1):46-53.

14. Rebollo, A., Pérez-Sala, D., Martínez-A. C. Bcl-2 differentially targets K-, N-, and H-Ras to mitochondria in IL-2 supplemented or deprived cells: implications in prevention of apoptosis. (1999) Oncogene18(35):4930-4939.

15. Joe, M.K., Kwon, H.S., Cojocaru, R., et al. Myocilin regulates cell proliferation and survival. (2014) J Biol Chem 289(14):10155-10167.

16. Rosseland, C.M., Wierod, L., Flinder, L.I., et al. Distinct functions of H-Ras and K-Ras in proliferation and survival of primary hepatocytes due to selective activation of ERK and PI3K. (2008) J Cell Physiol 215(3):818-826.

17. Copin, B., Brézin, A.P.,Valtot, F., et al. Apolipoprotein E-promoter single-nucleotide polymorphisms affect the phenotype of primary open-angle glaucoma and demonstrate interaction with the myocilin gene. (2002) Am J Hum Genet 70(6):1575-1581.

18. Zhang, X., Clark, A.F.,Yorio, T. Regulation of glucocorticoid responsiveness in glaucomatous trabecular meshwork cells by glucocorticoid receptor-beta. (2005) Invest Ophthalmol Vis Sci 46(12):4607-4616.

19. Mookherjee, S., Acharya, M., Banerjee, D., et al. Molecular Basis for Involvement of CYP1B1 in MYOC Upregulation and Its Potential Implication in Glaucoma Pathogenesis. (2012) PLoS ONE 7(9): e45077.

20. Beaumont, P.E., Kang, H.K. Clinical characteristics of retinal venous occlusions occurring at different sites. (2002) Br J Ophthalmol 86(5):572-580.

21. Fielding, P.E., Ishikawa, Y., Fielding, C.J. Apolipoprotein E me-

diates binding of normal very low density lipoprotein to heparin but is not required for high affinity receptor binding. (1989) J Biol Chem 264(21):12462-12466.

22. Goldwich, A., Scholz, M., Tamm, E.R. Myocilin promotes substrate adhesion, spreading and formation of focal contacts in podocytes and mesangial cells. (2009) Histochem Cell Biol 131(2):167-180.

23. Reeves, E.P., Lu H, Jacobs, H., et al Killing activity of neutrophils is mediated through activation of proteases by K+ flux. (2002) Nature 416(6878):291-297.

24. Pixley, R.A., Lin, Y., Isordia-Salas, I., et al. Fine mapping of the sequences in domain 5 of high molecular weight kininogen (HK) interacting with heparin and zinc. (2003) J Thromb Haemost 1(8):1791-1798.

25. Strittmatter, W.J., Roses, A.D. Apolipoprotein E and Alzheimer disease. (1995) Proc Natl Acad Sci U S A 92(11):4725-4727.

26. Bofinger, D.P., Feng, L., Chi, L.H, et al. Effect of TCDD exposure on CYP1A1 and CYP1B1 expression in explant cultures of human endometrium. (2001) Toxicol Sci 62(2):299-314.

27. Dong, H., Shertzer, H.G., Genter, M.B., Mitochondrial targeting of mouse NQO1 and CYP1B1 proteins. (2013) Biochem Biophys Res Coummun 435(4):727-732.

28. Torrado, M., Trivedi, R., Zinovieva, R., et al. Optimedin: a novel olfactomedin-related protein that interacts with myocilin. (2002) Hum Mol Genet 11(11):1291-1301.

29. Sohn, S., Joe, M.K., Kim, T.E., et al. Dual localization of wild-type myocilin in the endoplasmic reticulum and extracellular compartment likely occurs due to its incomplete secretion. (2009) Mol Vis 15:545-556.

30. Lemansky, P., Smolenova, E., Wrocklage, C., et al. Neutrophil elastase is associated with serglycin on its way to lysosomes in U937 cells. (2007) Cell Immunol 246(1):1-7.

31. Scotti, E., Calamai, M., Goulbourne, C.N., et al. IDOL stimulates clathrin-independent endocytosis and multivesicular body-mediated lysosomal degradation of the low-density lipoprotein receptor. (2013) Mol Cell Biol 33(8):1503-1514.

32. Shin, D.S., Kim, H.N., Shin, K.D., et al. Cryptotanshinone inhibits constitutive signal transducer and activator of transcription 3 function through blocking the dimerization in DU145 prostate cancer cells. (2009) Cancer Res 69(1): 193-202.