International Journal of Life science & Pharma Research

IN SILICO ANALYSIS OF PROTEIN-PROTEIN INTERACTION NETWORK OF HUTCHINSON GILFORD PROGERIA SYNDROME

SAPANA SINGH YADAV* AND USHA CHOUHAN

Department of Bioinformatics, Maulana Azad National Institute of Technology Bhopal – 462003, India

ABSTRACT

HGPS is a rare genetic disorder, caused by mutations in genes, encoding proteins of the nuclear lamina. Analysis of protein interaction network in the cell would be the key to understand how complex processes, lead to diseases. Protein-protein interaction network (PPIN) analysis provides the possibility to quantify the hub proteins in large networks as well as their interacting partners. A comprehensive genes/proteins dataset related to HGPS is created by analysing public proteomic data and text mining of scientific literature. From this dataset the associated PPI network is acquired to understand the relationships between topology and functionality of the PPI network. The extended network of seed proteins network consisted of 128 nodes connected via 376 edges (Fusion) and 127 nodes connected via 377 edges (Coexpression), targeted for analysis. The backbone network derived from giant network with high BC proteins presents a clear and visual overview which shows closely related proteins of HGPS and the crosstalk between them. Proteins with high BC and large degree have been identified as backbone network of disease. LMNA with highest BC and CC located in the centre of the network. Finally, the robustness of central proteins and accuracy of backbone are validated by 127 test networks. Based on the network topological parameters such as degree, closeness centrality, betweenness centrality we conclude that integrated PPIN is centred on LMNA. Although finding of other interacting partners are strongly represented as novel drug targets for HGPS.

KEYWORDS: *PPIN, Closeness Centrality (CC), Betweenness Centrality(BC), Degree, HGPS*

INTRODUCTION

Hutchinson-Gilford Progeria Syndrome (HGPS) is a lethal congenital disorder, characterised by premature ageing in children, caused by a point mutation in the *lamin A* gene.[1-2]Although HGPS was first described by Jonathan Hutchinson[3] and then by Hastings Gilford,[4] more than a century ago, it was not until 2003 that the genetic basis of HGPS was uncovered.[2-5] Manifestations of HGPS typically appear before 24 months of age (HGPS Research database,www.HGPSresearch.org),[6] and include loss of subcutaneous fat, severe growth retardation, hair loss, bone deformations, osteoporosis, delayed dentition, joint stiffness, hip-dislocations, sclerodermatous areas, and progressive arteriosclerosis. HGPS patients have an aged appearance, and in the final stages of disease, most children have hypertension, angina, and dilated hearts because of atherosclerotic heart disease. Children with HGPS generally die of myocardial infarction or cerebrovascular accident at an average age of thirteen years.[7] Systems approaches aim to develop an understanding of the inter-relationships between proteins, metabolites or other molecules.[8] Modern high-throughput techniques, taking measurements on a system-wide level, are well suited to the global analysis and modelling of networks for different diseases.[9-10] In comparison to wet lab techniques, computational methods have the potential to reduce noise and systematic errors.[11] Protein complexes are remarkable for understanding principles of cellular organization and function.[8] High throughput experimental techniques have generated a large amount of protein interactions, which make it doable to uncover protein complexes from protein protein interaction networks.[12-13]A PPI network (PPIN) can be modelled as an undirected graph, where vertices stand for proteins and edges represent interactions between proteins.[14] Protein complexes are set of

proteins that interact with one another, typically dense subgraphs in PPI networks.[14-15] To reveal the significance of the HGPS disease, insilico based methodology has been used to identify the key proteins and their interactors.[16-17] The integration of proteins interface structure into interaction graph models gives a better explanation of hub proteins, and builds up the relationship between the role of the hubs in the cell and their topological properties.[18-20] In this study, the interactions among the proteins have been implemented to produce and analyse a giant network by the topological analysis of the PPIN derived from the genes/proteins related to HGPS.[3-6] Different bioinformatics tools related to the proposed methodology have been implemented to construct the PPI network of candidate genes and analyzed the topological properties like degree, betweenness centrality (BC) and closeness centrality (CC).[20-21]

METHOD

Research method used in this study mainly included five steps, first step: Extraction of candidate genes, second step: Construction of PPIN of the seed proteins, third step: Merging of all PPIN scanned from seed proteins, fourth step: Analysis of the giant PPIN according to topological properties, fifth step: Acquiring backbone network.

Extraction of the candidate gene

For the extraction of the candidate genes related to HGPS; the PolySearch text mining system and NCBI database have been considered.[22] PolySearch is a web-based text mining system for extracting relationships between human diseases, genes, mutations, drugs and metabolites, It can produce relevant information regarding to an individual query. The query type is 'Disease-Gene/Protein Association' and the query keyword is 'HGPS'. So using this tool we fetched 58 candidate genes associated with HGPS disease. To check the accuracy, we manually confirmed whether these genes are associated with HGPS, and shorted the genes on the basis of Z score value >0. Finally a total of fourteen candidate genes were obtained, Table 1.

Construction of PPI network of the seed proteins

To Identify PPI interactions of the seed proteins STRING database has been used.[23] Interactions in STRING are provided with a confidence score, and accessory information such as protein domains and 3D structures are made available, all within a stable and consistent identifier space.[23]Fusion and coexpression attributes have been fixed to construct the PPI network, which are appropriate to consider for analysis.[24]

Merging of all PPI network scanned from seed proteins

To merge all the interacting networks of seed proteins within a single network for visualization and analysis Cytoscape v3.0.1 has been used.[25] Merged network includes different clustering of the seed proteins which result in formation of distinguish networks, Figure 1, considering only one network with highest number of existing nodes and edges, which have maximum interactions among the seed proteins and termed as giant network. Giant network has been extracted from the merged network after omitting other small networks.

Table 1
The list of genes extracted from NCBI and PolySearch Text mining system database showing association with HGPS

SN	Symbol	Description
1	BANF1	Barrier To Autointegration Factor 1
2	C myc	Avian Myelocytomatosis
3	DDX12	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 12, pseudogene
4	ELN	Elastin
5	EMD	Emerin
6	ERCC1	excision repair cross-complementing rodent repair deficiency, complementation group 1
7	ERCC4	excision repair cross-complementing rodent repair deficiency, complementation group 4
8	ROBO3	Roundabout, Axon Guidance Receptor, Homolog 3
9	LMNA	lamin A/C
10	MMP20	Matrix Metallopeptidase 20
11	SIRT1	Sirtuin 1
12	SUN2	Sad1 and UNC84 domain containing 2

13	WRN	werner syndrome, RecQ helicase like
14	ZMPSTE24	zinc metallopeptidase STE24

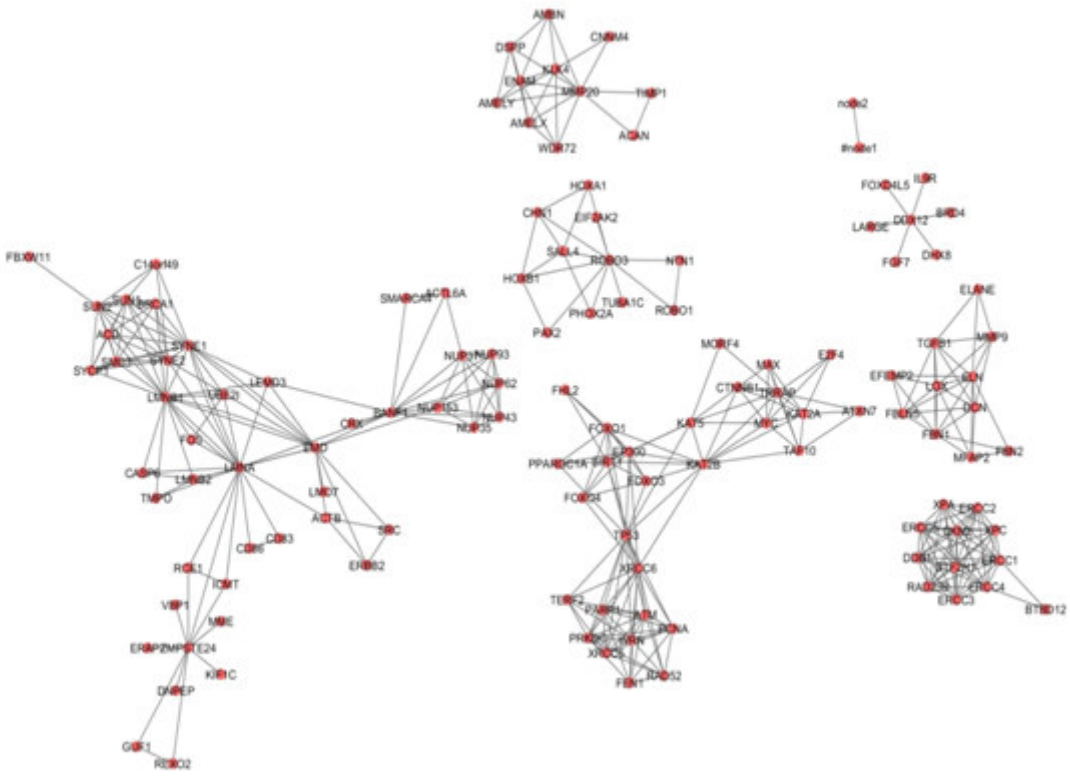


Figure 1
Overview of the extended network. Fusion attribute, 128 nodes and 376 edges, includes one giant network and six separated small networks

Analysis of the giant PPI network according to Topological Properties

PPIN of relevant disease represented by an undirected graph G(V, E), where V represents the set of vertices in the graph G and E represents the set of edges.[26] NetworkAnalyzer, used to compute various network parameters.[27] To predict and study the key nodes or hub proteins of the giant network topological parameters have been calculated. Therefore, for analyzing the giant network the degree, BC and CC values for each node have been calculated. That helps in finding the proteins of central positions in the network, as they can be highly important from a functional point of view too. In undirected networks, the node degree of a node *n* is the number of edges linked to *n*.[28-29] The number of links of a node was observed to follow a power law distribution, that is, the probability of a node having degree k is proportional to k−γ, and the distribution is independent of the number of nodes; hence these networks are called scale free. Scale-free networks have many nodes with small degrees and allow nodes with high degrees (hubs) with decreasing probability.[30] Betweenness measures how often nodes occur on the shortest paths between other nodes.[31] For a graph G(V, E), with n vertices, the betweenness centrality $C_B(v)$ a vertex v is defined in equation (1),

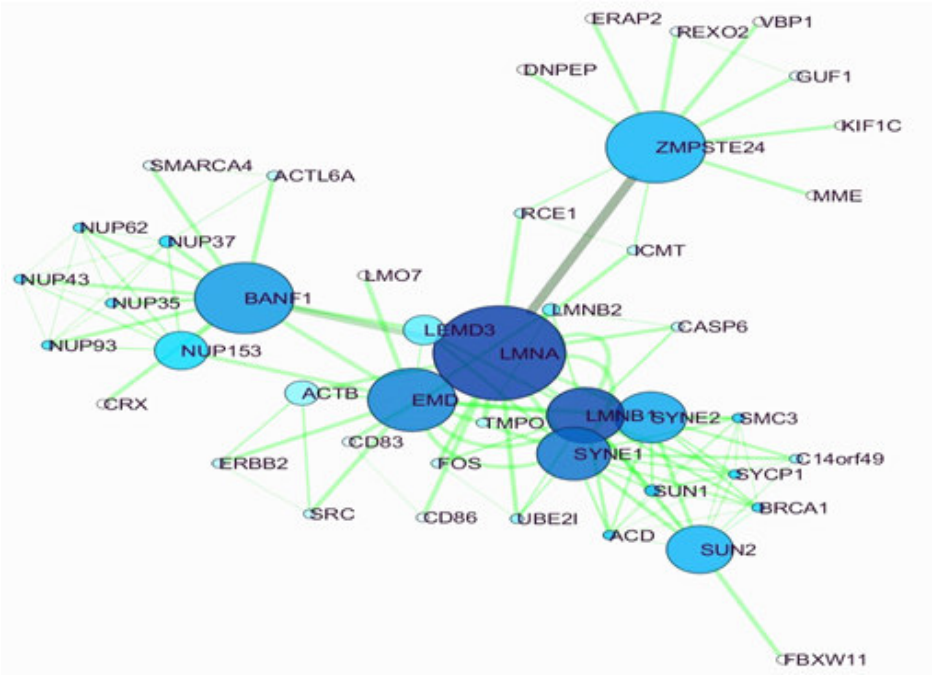
$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \qquad (1)$$

Where σ_{st} is the number of shortest paths from *s* to *t*, and $\sigma_{st}(v)$ is the number of shortest paths from *s* to *t* that passes through a vertex *v*.

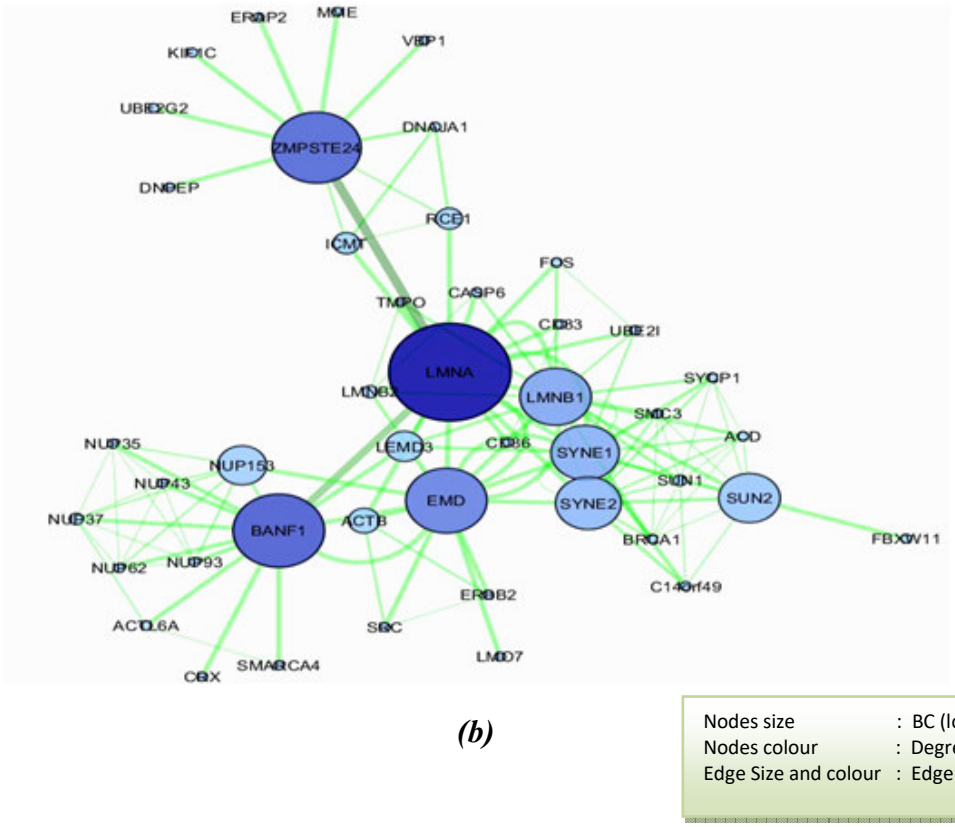
Closeness centrality[32] $C_c(n)$ of a node *n* is defined as the reciprocal of the average shortest path length and is computed as,

$$C_c(n) = \frac{1}{avg(L(n,m))} \qquad (2)$$

Where $L(n,m)$ is the length of the shortest path between two nodes n and m. The closeness centrality of each node is a number between 0 and 1. In PPIN the nodes with high degree defined as hub proteins and the nodes with high betweenness as bottleneck proteins.[19-20]



(a)



(b)

Figure 2

The topology of the giant network. The giant network extracted from the extended network is the biggest component in the extended network.(a) Fusion 45 nodes and 125 edges (b) Coexpression,45 nodes 132 edges

Table 2
The list of high BC nodes and their corresponding CC values
(a) Coexpression (b) Fusion

(a) ***(b)***

SN	NODE	BC	CC	SN	NODE	BC	CC
1	LMNA	0.53913997	0.611111	1	LMNA	0.53913997	0.61111111
2	BANF1	0.29134176	0.478261	2	ZMPSTE24	0.294926	0.44444444
3	ZMPSTE24	0.26920366	0.444444	3	BANF1	0.29134176	0.47826087
4	EMD	0.20767941	0.52381	4	EMD	0.20767941	0.52380952
5	LMNB1	0.1278452	0.505747	5	LMNB1	0.1278452	0.50574713
6	SYNE1	0.10202344	0.488889	6	SYNE1	0.10202344	0.48888889
7	SYNE2	0.08122821	0.478261	7	SYNE2	0.08122821	0.47826087
8	SUN2	0.04756242	0.369748	8	SUN2	0.04756242	0.3697479
9	NUP153	0.03009939	0.389381	9	NUP153	0.03009939	0.38938053
10	LEMD3	0.02118878	0.483516	10	LEMD3	0.02118878	0.48351648
11	ACTB	0.01603242	0.407407	11	ACTB	0.01603242	0.40740741
12	ICMT	0.01233263	0.419048	12	LMNB2	0.00334743	0.44
13	RCE1	0.01233263	0.419048	13	NUP37	0.00264271	0.34108527
14	LMNB2	0.00334743	0.44	14	SUN1	0.00210787	0.36666667
15	NUP37	0.00264271	0.341085	15	ACTL6A	5.29E-04	0.33082707

Table 3
The list of large Degree nodes and their CC values
(a) Coexpression (b) Fusion

(a) ***(b)***

SN	Node	Degree	CC	SN	Node	Degree	CC
1	LMNA	19	0.61111111	1	LMNB1	21	0.505747
2	LMNB1	18	0.50574713	2	LMNA	20	0.611111
3	SYNE1	15	0.48888889	3	SYNE1	18	0.488889
4	EMD	14	0.52380952	4	EMD	16	0.52381
5	BANF1	12	0.47826087	5	BANF1	13	0.478261
6	SYNE2	11	0.47826087	6	SYNE2	13	0.478261
7	ZMPSTE24	10	0.44444444	7	ZMPSTE24	10	0.444444
8	SUN2	10	0.3697479	8	SUN2	10	0.369748
9	SUN1	9	0.36666667	9	SUN1	9	0.366667
10	SMC3	8	0.36363636	10	SMC3	8	0.363636
11	ACD	8	0.36363636	11	ACD	8	0.363636
12	NUP153	7	0.38938053	12	NUP153	7	0.389381
13	NUP37	7	0.34108527	13	NUP37	7	0.341085
14	SYCP1	7	0.36065574	14	SYCP1	7	0.360656
15	BRCA1	7	0.36065574	15	BRCA1	7	0.360656

RESULTS

Giant Network

Fourteen numbers of candidate genes related to HGPS disease after using PolySearch Text mining tools were collected. Seed proteins have been generated from STRING. At the beginning step to construct the merged network, we fixed attributes, therefore, according to each individual attribute two different merged networks formed. First, fusion attribute- the merged network had been combination of seven different networks. LMNA, DDX12, SIRT1, ROBO3, ELN, MMP20, ERCC1 are the seed proteins as well as play the central role in each seven sub networks. The merged network consists of 128 nodes and 376 edges. These nodes are distributed in seven different clusters according to interaction possibility, so seven distinct sub networks had been formed. The large network among them consists of 45 nodes and 125 edges

taken as giant network. Similarly, considering the coexpression attribute the merged network consists of 127 nodes and 377 edges, Figure 2(a,b) It is notified that in both cases previously said seed proteins have the key role in each sub network. Similar to fusion attribute in case of coexpression attribute LMNA is the key protein in the giant network. The giant network consists of 45 nodes and 132 edges according to coexpression attributes.

Key nodes in the PPIN: Backbone Network

To study the key nodes of giant network according to BC, CC and degree for each node has to be measured and comparison can be made according to values for different attributes. Topological statistics of network calculated with NetWorkAnalyzer, among them highest 15 nodes with corresponding BC, CC and degree extracted, are summarized in Table 2 and Table 3. To discriminate the nodes with high BC value fixed the threshold at 15% of the total node set of the network, i.e. 7 proteins with high BC value have been chosen. Among these seven proteins LMNA at highest with BC value 0.56, ZMPSTE24, BANF1, EMD, LMNB1, SYNE1, SYNE2 other proteins with high BC value, Table 4. The links in between these proteins are considered to construct backbone of the network, Figure 3. Concisely, it is possible to measure the number of shortest paths among those nodes which have high BC value rather than considering all nodes of the network.

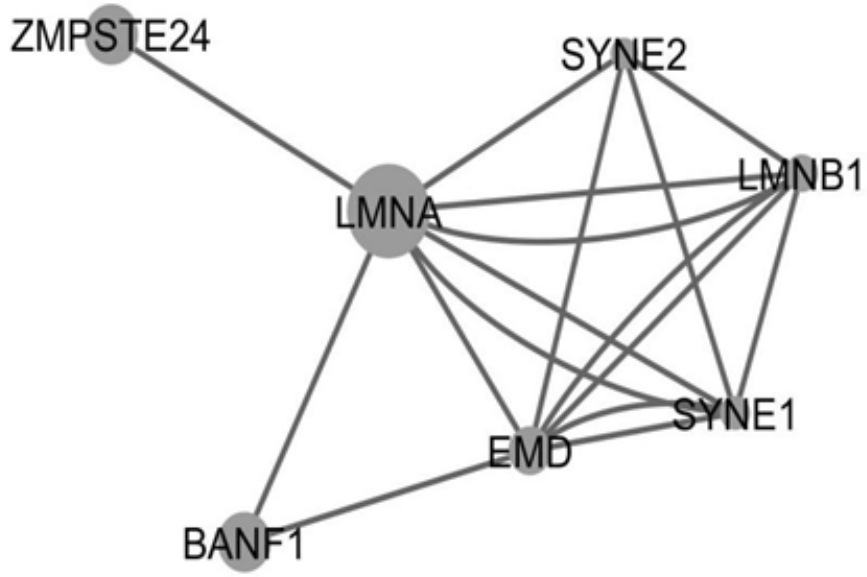


Figure 3

The topology of the backbone network. The backbone network consists of 7 nodes with high BC value, size of nodes corresponds to their BC values.

Sub network consisting of all shortest paths

To analyze a sub network in which all related proteins to the HGPS disease to be connected directly or indirectly with minimum number of nodes, consider the network that consists of those nodes which implied the shortest paths between every pair of the seed proteins. But this step is not compulsory if the giant network consists of all such shortest path. From the construction of the sub network it can be concluded that LMNA has the highest BC value in comparison with all other candidate gene of the HGPS disease. This sub network consists of 65 nodes and 2 proteins which are neither having high BC value nor seed proteins. This result is summarized according to the fusion parameter. Similarly the result according to the other parameter coexpression was also finding out.

The Robustness of the backbone network and LMNA as a Central Protein

If we consider the backbone of any network with nodes along with high BC values then validation of the backbone network and the LMNA as central protein is to be verified. For this purpose we constructed different testing network using some number of genes from total genes list, considering as initial seeds. Now omitting genes randomly from the range between 1 to 7. In this random sampling one can make different combination of omitting genes from the list of seed proteins. When the number of omitting genes is higher or equal to three then we consider two genes randomly and in each case LMNA is considered as fixed gene which have to be omitted every time and only 20 times

randomly selected genes are taken to examine the test network. So therefore almost 127 test networks had been constructed and BC values calculated accordingly. Among the total of 127 test network, the number of frequency of LMNA in test network is 89. The frequency of LMNA has to be calculated with the largest BC value and the accuracy of the back bone network has to be measured too, to validate the LMNA as a central protein of the general interaction network as well as back bone network. The accuracy of the back bone network is 0.78198, Table 5. It is examined that whenever the number of omitting genes was larger than 3 then the accuracy of back bone network and frequency of the LMNA decreased continuously.

Table 4
The backbone network consists of 7 nodes with high BC value Fusion

SN	Node	BC
1	LMNA	0.53913997
2	ZMPSTE24	0.294926
3	BANF1	0.29134176
4	EMD	0.20767941
5	LMNB1	0.1278452
6	SYNE1	0.10202344
7	SYNE2	0.08122821

Table 5
Frequency of nodes with the largest BC value and accuracy of backbone in the 127 test networks

Number of omitted genes	Frequency of nodes with the largest BC value in the test networks							Accuracy of the backbone	Number of the test networks
	LMNA	ZMPSTE24	BANF1	EMD	SYN1	SYN2	LMNB1		
1	14	0	0	0	0	0	0	0.88277	14
2	13	0	0	0	0	0	0	0.84548	13
3	14	4	1	0	1	0	0	0.76858	20
4	13	4	2	0	0	0	1	0.74596	20
5	12	5	1	1	0	0	1	0.74485	20
6	11	5	1	1	0	1	1	0.74365	20
7	12	4	2	0	0	0	2	0.74258	20
Summary	**89**	**22**	**7**	**2**	**1**	**1**	**5**	**0.78198**	**127**

Table 6
Comparative network statics for Fusion and Coexpression

SN	Network statics	Fusion	Coexpression
1	Clustering Coefficient	1	0.638
2	Network diameter	5	5
3	Network radius	3	3
4	Network centralization	0.276	0.275
5	Shortest paths	1980	1992
6	Characteristic path length	2.690	2.689
7	Avg. No.of neighbours	5.378	5.422
8	Number of nodes	45	45
9	Network density	0.122	0.123
10	Network heterogeneity	0.768	0.759

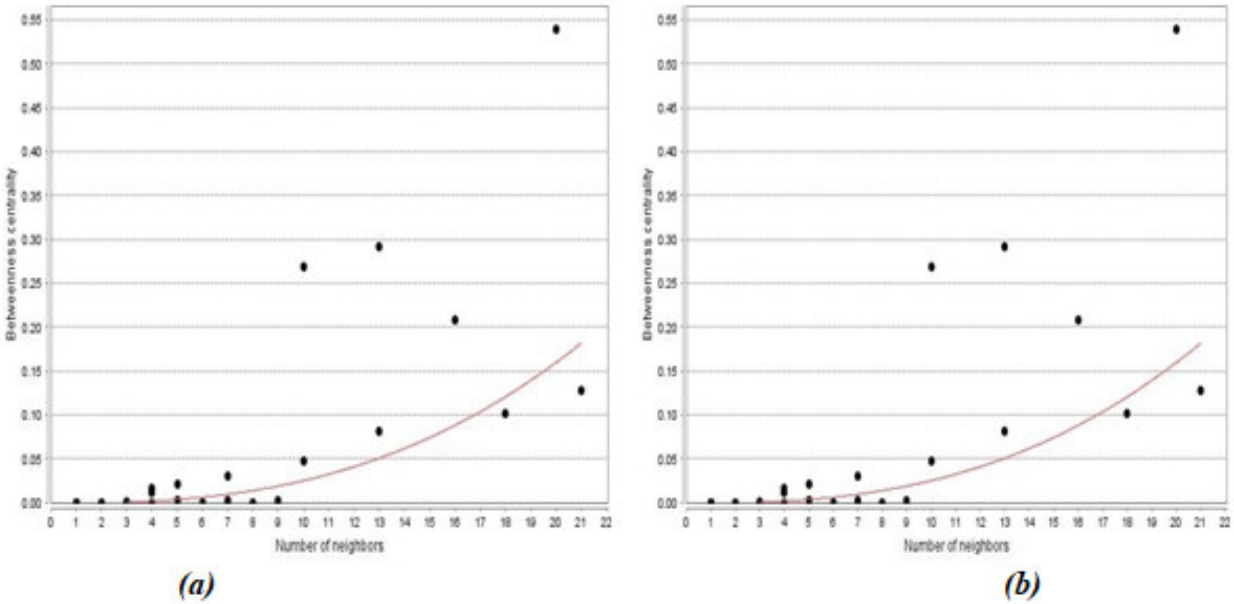


Figure 4
Betweenness centrality of the network with a fitted line
(a) Fusion (b) Coexpression

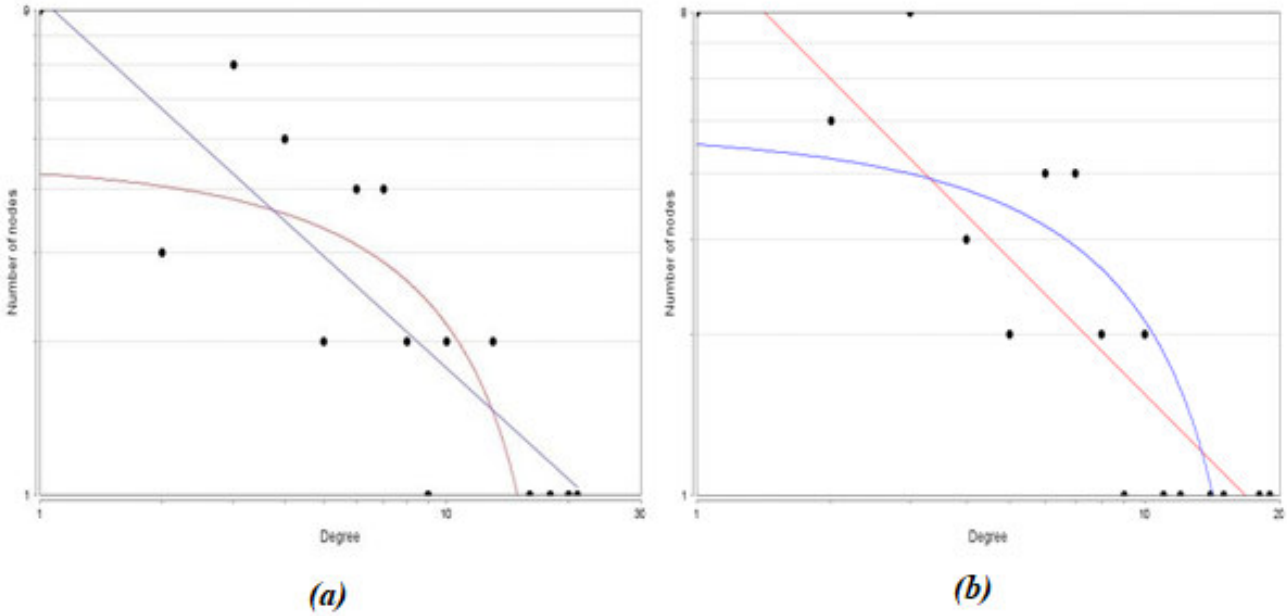


Figure 5
Node Degree distribution of the network with a fitted power law, R-squared value reported is the R-squared value for the fitted line on logarithmized data.
(a) Fusion (b) Coexpression

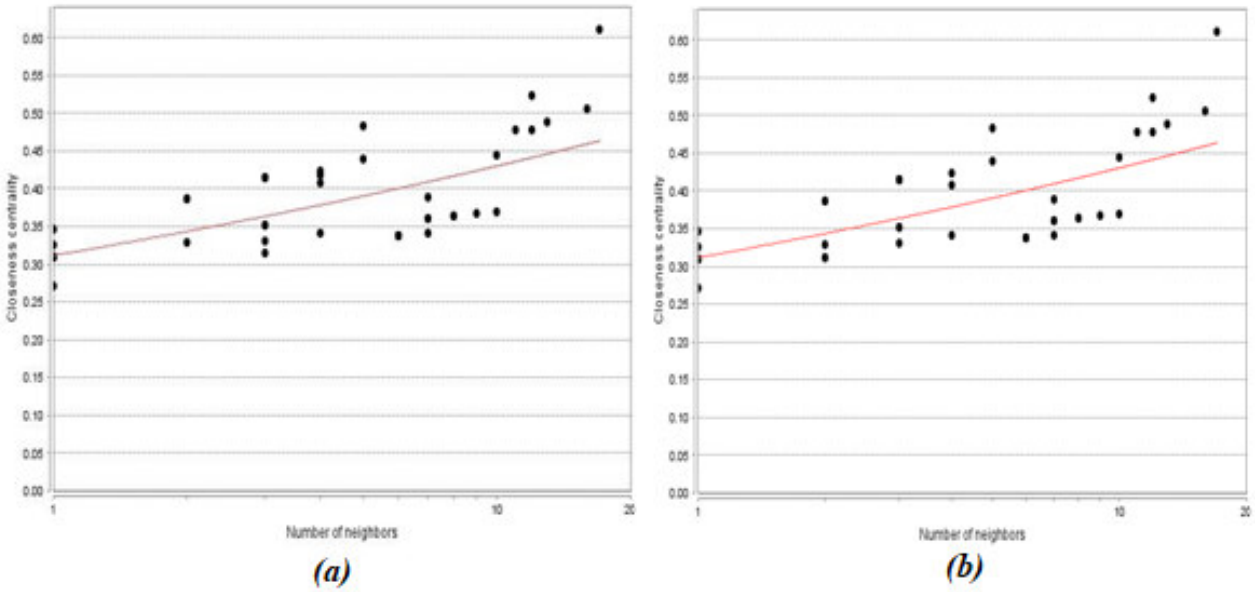


Figure 6
Closeness centrality of the network with a fitted line.
(a) Fusion (b) Coexpression

DISCUSSION

The purpose of this study is to find out the essential key proteins related to HGPS. Insilico approaches of bioinformatics had been taken to validate the literature about the list of key proteins mainly involved in the disease network spreading. So from this model we suggest that the central proteins in the PPI network can be even more important in systems based medicine and drug design than either the hubs or the disease proteins themselves. Here the motto is to identify the hub of proteins which are functionally expressed in the case of disease expression. Those hub proteins of the PPI network may be of lower topological importance in the groups of disease proteins but they are involved in the disease mechanism from the initial expression level to functionally expressed level and most importantly such hubs of proteins play the role of mediator to activate other genes of the network and indirectly they have huge effect on the construction of the merged network. By quantitatively identifying, the most important mediator proteins which are not present in the hubs of the proteins also play key roles in maintaining communication between disease proteins or genes. From the topological analysis it is clear that LMNA is center key protein as well as the involvement of other key proteins BANF1, ZMPSTE24, EMD, LMNB1, SYN1, SYN2 in the different stage of the disease means that without interactions with these proteins the disease interaction network is not possible to construct. In this attempt a study was also made on the comparative analysis of the network according to two most important attributes to understand how the attributes can make an effect on the network analysis in case of our experiment. In following Table 6 the comparison between fusion and coexpression has been summarized. The results of all the Parameters have approximately same value, only shortest path in case of coexpression is little bit high which does not affect the other parameter like BC value, CC value, clustering coefficient,etc. Graphical results of different topological parameters shown in Figure 4 (a,b), explains that the highest betweenness centrality in the giant network is approximate 0.54 and number of nodes is 20. This implies, the node having the highest betweenness value also has the highest number of neighbors which signifies evidences of the key node of the network. If we compare the second highest beetweenness value of the network, it is 0.29 and consists of around thirteen neighbors. Therefore the node having the first position in both cases of BC value and neighborhood, proving better candidature for the key role in extended merged giant network. NetworkAnalyzer can fit a power law to some topological parameters and follow the least squares method,[33] and only points with positive coordinate values are considered for the fit, gives the correlation between the given data points and the corresponding points on the fitted curve. In addition, the R-squared value (also known as coefficient of determination) is reported. This coefficient gives the proportion of variability in a data set, which is explained by a fitted linear model.

Therefore, the R-squared value is computed on logarithmized data, where the power-law curve: $y = \beta x\alpha$ is transformed into linear model: $\ln y = \ln\beta + \alpha \ln x$., here correlation between the data points and corresponding points on the line is approximately 0.682 and 0.670, R-squared value is 0.392 and 0.389 respectively for fusion and coexpression. Figure 5 (a,b) show a, graphical representation of the number of nodes in a giant network, according to degrees, the distribution of those nodes which are following minimum number of connectivity i.e. nodes are connected by at least one edge. It was identified that when the number of nodes are 08 then the degree of such nodes is 04. It was also, observed that in some cases where the number of degrees was high, the number of nodes were less. This implies such nodes are not part of giant network and they made subnetwork which contains less nodes. Therefore the connectivity is high, but the node is less. NetworkAnalyzer provides another useful feature - fitting a line on the data points of some complex parameters. The method applied is the least squares method for linear regression.[33] Fitting a line can be used to identify linear dependencies between the values of the x and y coordinates in a complex parameter. The fitted line on degree, having correlation between the data points and corresponding points on the line is approximately 0.704 and 0.627, R-squared value is 0.386 and 0.394 respectively for fusion and coexpression. Figure 6 (a,b), explains the value of closeness centrality of each node of the giant network, according to the number of neighbors. Clearly, it shows that only single node consists of highest CC value which is 0.61 approximate worth having 18 neighbors and graph also fitted to power law having correlation between data points and corresponding point on the line is approximately 0.218 and 0.223, R-squared value is 0.560 and 0.568.

CONCLUSION

In the present study, a comprehensive initial dataset of genes statistically related to HGPS and a further expansion through the construction of related PPIN were created. Relationships between interacting proteins according to topological properties were studied. It was shown that a protein or a hub of proteins can play an important role to interact with other proteins and also extend the PPI disease network. Again, it is possible to find out the key proteins, which are main mediator for disease networks. Identifying such hub of proteins can help to understand the mechanism of pathways it might also be possible to emphasize that they have high functional importance in the cell. Most of seed proteins associated with HGPS and their PPI neighbours are connected to a giant network, which is analyzed by using different centrality indexes for hubs detection. The findings suggest that HGPS disease mechanism and pathway is organized by an integrated PPI network centred on LAMIN gene product LMNA protein, while other proteins BANF1, ZMPSTE24, EMD, LMNB1, SYN1, SYN2 with high BC values predict their significant role in a network. The backbone network is robust against the changes of initial seed genes. The results may provide a basis for further experimental investigations to study PPI networks associated with HGPS and other relevant disease.

ACKNOWLEDGEMENTS

The authors are cordially thankful to the Madhya Pradesh Council of Science and Technology, Bhopal for providing financial support to carry out this work.

Contribution of authors

We declare that this work was done by the author Sapana Singh Yadav and Dr. Usha Chouhan and all liabilities pertaining to claims relating to the content of this article will be borne by the authors. Each author contributed equally.

CONFLICT OF INTEREST

Conflicts of interest declared none.

REFERENCES

1. Eriksson M, Brown WT, Gordon LB, Glynn MW, Singer J, Scott L, Erdos MR, Robbins CM, Moses TY, Berglund P, Dutra A. Recurrent de novo point mutations in lamin A cause Hutchinson–Gilford progeria syndrome. Nature. 2003 May 15;423(6937):293-8.
2. Scaffidi P, Misteli T. Reversal of the cellular phenotype in the premature aging disease Hutchinson-Gilford progeria syndrome. Nature medicine. 2005 Apr;11(4):440.
3. Goldman RD, Shumaker DK, Erdos MR, Eriksson M, Goldman AE, Gordon LB, Gruenbaum Y, Khuon S, Mendez M, Varga R, Collins FS. Accumulation of mutant laminA causes progressive changes in nuclear

architecture in Hutchinson–Gilford progeria syndrome. Proceedings of the National Academy of Sciences of the United States of America. 2004 Jun 15;101(24):8963-8.
4. Uitto J. Searching for clues to premature aging. Trends in Endocrinology & Metabolism. 2002 May 1;13(4):140-1.
5. Schreiber KH, Kennedy BK. When lamins go bad: nuclear structure and disease. Cell. 2013 Mar 14;152(6):1365-75.
6. Davidson PM, Lammerding J. Broken nuclei–lamins, nuclear mechanics, and disease. Trends in cell biology. 2014 Apr 30;24(4):247-56.
7. De Sandre-Giovannoli A, Bernard R, Cau P, Navarro C, Amiel J, Boccaccio I, Lyonnet S, Stewart CL, Munnich A, Le Merrer M, Lévy N. Lamin a truncation in Hutchinson-Gilford progeria. Science. 2003 Jun 27;300(5628):2055-EOF.
8. Csoka AB, Cao H, Sammak PJ, Constantinescu D, Schatten GP, Hegele RA. Novel lamin A/C gene (LMNA) mutations in atypical progeroid syndromes. Journal of Medical Genetics. 2004 Apr 1;41(4):304-8.
9. Hutchinson J. Congenital absence of hair and mammary glands with atrophic condition of the skin and its appendages, in a boy whose mother had been almost wholly bald from alopecia areata from the age of six. Medico-chirurgical transactions. 1886;69:473.
10. Gilford H. On a condition of mixed premature and immature development. Medico-chirurgical transactions. 1897;80:17.
11. Bruggeman FJ, Westerhoff HV. The nature of systems biology. TRENDS in Microbiology. 2007 Jan 31;15(1):45-50.
12. Alm E, Arkin A. Biological networks. Current Opinion in Structural Biology. 2003 13:193–202
13. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nature reviews. Genetics. 2004 Feb 1;5(2):101.
14. Mishra S, Mishra D. An Overview Of Biological Networks: Mechanisms, Methodologies And Applications, Int J Pharm Bio Sci 2016 July ; 7(3): (B) 979 – 988
15. Ran J, Li H, Fu J, Liu L, Xing Y, Li X, Shen H, Chen Y, Jiang X, Li Y, Li H. Construction and analysis of the protein-protein interaction network related to essential hypertension. BMC systems biology. 2013 Apr 12;7(1):32.
16. LaCount DJ, Vignali M, Chettier R, Phansalkar A. A protein interaction network of the malaria parasite Plasmodium falciparum. Nature. 2005 Nov 3;438(7064):103.
17. Gilchrist MA, Salter LA, Wagner A. A statistical framework for combining and interpreting proteomic datasets. Bioinformatics. 2004 Jan 22;20(5):689-700.
18. Raman K. Construction and analysis of protein–protein interaction networks. Automated experimentation. 2010 Feb 15;2(1):2.
19. Zhang Y, Lin H, Yang Z, Wang J. Integrating experimental and literature protein-protein interaction data for protein complex prediction. BMC genomics. 2015 Jan 21;16(2):S4.
20. Ryan DP, Matthews JM. Protein–protein interactions in human disease. Current opinion in structural biology. 2005 Aug 31;15(4):441-6.
21. Doncheva NT, Assenov Y, Domingues FS, Albrecht M. Topological analysis and interactive visualization of biological networks and protein structures. Nature protocols. 2012 Apr 1;7(4):670.
22. Xu J, Li Y. Discovering disease-genes by topological features in human protein–protein interaction network. Bioinformatics. 2006 Sep 5;22(22):2800-5.
23. Assenov Y. Topological Analysis of Biological Networks. 2006 (Doctoral dissertation, Max Planck Institute for Informatics).
24. Cheng D, Knox C, Young N, Stothard P, Damaraju S, Wishart DS. PolySearch: a web-based text mining system for extracting relationships between human diseases, genes, mutations, drugs and metabolites. Nucleic acids research. 2008 May 16;36(suppl_2):W399-405.
25. Von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, Foglierini M, Jouffre N, Huynen MA, Bork P. STRING: known and predicted protein–protein associations, integrated and transferred across organisms. Nucleic acids research. 2005 Jan 1;33(suppl_1):D433-7.
26. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research. 2003 Nov 1;13(11):2498-504.
27. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-

Campilo I, Creech M, Gross B, Hanspers K. Integration of biological networks and gene expression data using Cytoscape. Nature protocols. 2007;2(10):2366.

28. Dwivedi A K, Chouhan U. Simulated Annealing Model For Reticulate Evolution In Molecular Sequences, Int J Pharm Bio Sci 2013 Oct; 4(4): (B) 497 - 503
29. Barabási AL, Albert R. Emergence of scaling in random networks. science. 1999 Oct 15;286(5439):509-12.
30. Brandes U. A faster algorithm for betweenness centrality. Journal of mathematical sociology. 2001 Jun 1;25(2):163-77.
31. Newman ME. A measure of betweenness centrality based on random walks. Social networks. 2005 Jan 31;27(1):39-54.
32. Goñi J, Esteban FJ, de Mendizábal NV, Sepulcre J, Ardanza-Trevijano S, Agirrezabal I, Villoslada P. A computational analysis of protein-protein interaction networks in neurodegenerative diseases. BMC systems biology. 2008 Jun 20;2(1):52
33. Weisstein EW. Chebyshev polynomial of the first kind- A Wolfram Web Resource. (http://mathworld.wolfram.com/LeastSquaresFittingPowerLaw.html)