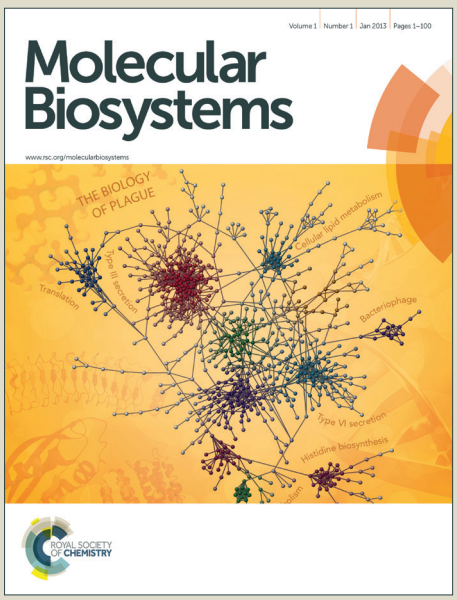
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Prediction of candidate genes for cervix related cancer through gene ontology and graph theoretical approach

V Hindumathi[1, 2], T Kranthi[1], S B Rao[1] and P Manimaran[1, $]

[1]C R Rao Advanced Institute of Mathematics, Statistics, and Computer Science, University of Hyderabad Campus, Prof. C R Rao Road, Gachibowli, Hyderabad -500046, India.

[2]Department of Plant Molecular Biology & Bioinformatics, Tamil Nadu Agricultural University, Coimbatore - 641003, India.

[$]**Corresponding author**

Dr. P Manimaran

Assistant Professor,

C R Rao Advanced Institute of Mathematics, Statistics and Computer Science,

University of Hyderabad Campus,

Prof. C R Rao Road, Gachibowli,

Hyderabad – 500046, INDIA.

Email: pa.manimaran@gmail.com

Mobile: +91-9989952362

Abstract

With rapidly changing technology, prediction of candidate genes has become an indispensable task in recent years mainly in the field of biological research. The empirical methods for candidate gene prioritization that succors to explore the potential pathway between genetic determinants and complex diseases are highly cumbersome and labor intensive. In such a scenario predicting potential targets for a disease state through insilico approaches are of researcher's interest. The prodigious availability of protein interaction data coupled with gene annotation renders an ease in accurate determination of disease specific candidate genes. In our work we have prioritized the cervix related cancer candidate genes by employing Csaba Ortutay and his co-workers approach of identifying the candidate genes through graph theoretical centrality measures and gene ontology. With the advantage of the human protein interaction data, cervical cancer gene sets and the ontological terms, we were able to predict 15 novel candidates for cervical carcinogenesis. The disease relevance of the anticipated candidate genes was corroborated through literature survey. Also presence of the drugs for these candidates was detected through Therapeutic Target Database (TTD) and DrugMap Central (DMC) which affirms that they may be endowed as a potential drug targets for cervical cancer.

Keywords: Graph theory, Gene ontology, Candidate genes, Cervical cancer.

Introduction

Pertaining to changing lifestyle, people are facing major life threatening disease like diabetes, cancer, hyper tension, heart disease and stroke [1]. These diseases take too long to develop but once they have developed in the body it is difficult to cure them. Among these, cancer considered as a genetic disease is controlled by multiple genes, and leads to unregulated cell growth where the tumors have taken the overall control of the body [2]. Tumors are biased accumulation of proliferating complex tissues with diverse cell types that are involved in heterotypic interactions with one another [3, 4]. A difference in the type of cancer/neoplasm is the result of various complex genetic and epigenetic events. Cancers can be classified into 200 different types of which third major death-causing tumors among women is the cervical cancer. Carcinoma of the cervix is a sexually transmitted disease caused by the human papillomavirus (HPV) infection that is formed on the squamous cells of the cervix [5].

A case study in cervical cancer through experimental analysis affirms that the differentially over expressed proteins were identified to be the novel gene for cervical cancer which were validated through Immunohistochemistry procedure. The disadvantage of the method lies in the arena that the novel genes identified had to be validated in a larger series of samples which is time consuming and highly challenging [6]. Likewise, the gene dosage and expression profiling techniques, and other insilico approaches can also be adopted to predict the candidates for cancer state but it requires prolonged time period to end up with a conclusion. Moreover cervical cancer is the outcome of involvement of several genes with low-to-moderate effects therefore it is more desirable to study the multilocus models and potential interactions between genes for disease gene prioritization [7]. The mechanisms and molecular architecture underlying various cancers

including cervical carcinoma can be better understood through the identification of other potential causal/susceptibility genes [2]. But the identification of such genesis a major bottleneck for the reason that genetics of cancer is still not well understood.

The disease causing genes are consistently explored since decades but it is still a quest to find the candidate genes underlying a specific disorder. Experimental studies such as linkage studies, gene expression analysis and genome wide classification studies have been found to be successful in identifying the high relative risk genes for a specific disease [8]. But the aforementioned methods have failed drastically in prioritizing the genes responsible for complex diseases. In this scenario, candidate gene approaches were found to be fruitful in identifying the risk variants associated with various diseases of interest such as dementia, cancer, diabetes, asthma, and hypertension [9-13]. Candidate genes are nothing but the genes with known biological function, suspected to be directly or indirectly involved in contributing to the aetiology of the disease [14]. The candidate gene approach is ubiquitously an imperative task that focuses on gene-disease research, biomarkers and drug target selection and has been proven to be powerful in studying the genetic architecture of complex traits and also an economical method for direct gene discovery [15]. This method has gained a considerable edge over the above mentioned approaches in terms of its quickness, simplicity, inexpensiveness, directness, high sensitivity for detecting the genes with small effect, and perfect plasticity in the practical application [16].

Recent advances in high throughput technologies paved a successful path for the candidate gene approaches. Experimental methods such as gene expression profiling, next generation sequencing, gene wide association studies, CHIP-seq, genetical linkage association studies generate candidate genes, [17-19]. The scarcity of disease associated families for linkage

analysis, large genomic regions, hindrances in identification of disease locus, lack of definitive functional conformation of the target gene and the labor insensitivity of the experimental methods urges for the development of various high end insilico approaches for disease gene prioritization.

In such circumstances, a number of insilico strategies have been consequently developed for candidate identification in various fields such as agriculture, biomedical, finance etc. These computer simulated methods have been grouped as ontology, computation and integration based candidate gene identification approaches. The ontology based approaches relies on the availability of annotated gene functional information on internet whereas the computation based approach prioritizes the genes through a computational framework utilizing the web resource based data sets [15]. Some of the computational methods include data mining analysis, Hidden Markov analysis, machine learning, kernel-based data fusion analysis, similarity- based method etc [20-24]. The integrated approach pools the information from various sources such as experimental data, web resource based data and many other features of protein-DNA interactions, molecular module, Protein-protein interactions, path way and gene regulatory networks etc [25-29]. Some of the computational tools that are publicly available online for prioritizing the candidate genes are SUSPECTS, POCUS, G2D, GFSST, ENDEAVOUR [30-34].The candidate gene approach backed by completed genome sequence of human and model organisms aids to dissect and identify genetic risk factors for cervical cancer [35]. But there exists only a limited number of platforms specialized for cancer gene identification which were proven to be less successful.

Most of the insilico candidate gene identification methods rely on the ontology based annotation approach which is nothing but the association of the biological phrases and specific genes. Gene

ontology encapsulates the known relation between biological terms and the genes that occur in these terms. This mode of action benefits the biologists to make inference considering cluster of genes rather than a single gene. The terms that are employed in gene ontology annotations are biological process, molecular function and cellular components. The biological process defines the biological phenomena affecting the state of an organism while the molecular function is specific to carry out the function by a gene product and the cellular component is concerned within the cell wherein a gene acts [36].

The problem ensued with the ontology based approaches is that only two thirds of human genes are being annotated and the rest of the fraction yet to be characterized [37]. With the tremendous escalation of human protein interaction data, the entanglement of the above techniques can be conquered through protein–protein interaction networks (PPINs) [38, 39]. Drastic changes that took place over several decades in the field of biological research towards massively parallel techniques creates new insight in this arena but creates problem in formulating meaningful information out of the generated data. These data could be expressed in the form of networks which provide structural annotation, where the nodes are proteins, linked by edges which are nothing but the interactions. Proteins are the representatives of the biological networks and they are realized only if the relationship between essentiality and topological properties such as the degree distribution, clustering coefficients, centrality measures, and community structures of the network are studied [40-42]. Of all the properties graph centrality measures aid in identification nodes that are functionally crucial in the network by ranking elements of a network. Different graph centrality measures such as vulnerability, closeness, centroid values, shortest-path betweenness, current-flow betweenness, and Eigen vector can be computed for every node in the interactome and rank them according to their scores which further aids in establishing the

properties of protein interaction network [43-46]. Thus the analysis of PPINs which are scale free in nature comforts the annotation of the uncharacterized genes facilitating the perception of disease mechanisms and thereby succors for disease gene prioritization [47]. However, even the network based approaches encounter certain limitations in terms of quality and availability of interaction data, missing interactions, false positives etc. The integration of both functional annotations (ontology approach) and network based topological parameters generates the information for protein functions, processes, localization and there by providing a more reliable approach for identification of candidate genes [48].

Thus the integrative computational approaches may be anticipated as the powerful tools for candidate gene identification, contributing to a major breakthrough in the field of cancer research. Protein-protein interaction networks and their properties provide valuable information to understand and analyze the mechanisms of disease particularly cancer [49-54]. The gene ontology terms facilitate the systematic annotation of the genes and thereby elucidate their biological relevance with the experimental results. Csaba Ortutay and his co-workers have already contributed a novel method for identifying candidate genes by consolidation of gene ontology and network based approaches employing only three graph centrality measures [55]. Our work, directs attention towards predicting candidate genes for cervical cancer through the same approach with the human protein interaction network, cervical genes and gene ontology terms, but with six different graph centrality measures. The advantage of using six different centrality measures is that each of them scores the proteins in an interactome based on different formalism/concept so that there exists a less chance of missing the biologically significant ones.

In our work we have utilized the gene ontology and network integrative approach of Csaba Ortutay et al which drastically reduces the time involved and efficiently predicts the potential

cervical cancer candidate genes with the availability of function, processes and localization information which is highly imperative in any cancer phenomenon. To find the genes that aid in the cervical cancer the protein interactome of all the cancer genes was constructed which resulted in human cancer gene network. A set of experimentally compiled cervical cancer genes is verified through network and gene ontology enrichment approaches. The above study on the cervical cancer furnished 15 novel genes which could be successful potential targets for drug discovery. These 15 genes may have a major role in either creating or causing the carcinogenic tumor in the cervix of women. The strategy of our work is shown in the graphical abstract **Figure-1**.

Materials and methods

Data collection and construction of human cancer gene network

The human protein interaction data was obtained from Human Integrated Protein-Protein Interaction rEference (HIPPIE) database [56]. The main purpose of using HIPPIE dataset is it focuses on likely true Protein-Protein Interaction (PPI) set by generating sub networks around proteins of interest. HIPPIE is an integrated set of human protein interaction data that is constructed according to experimental evidences.The database contains 11,468 proteins with 70,401 human PPIs which are obtained in combination with all the major PPI datasets such as HPRD, MINT, DIP etc [57-59].

The cancer genes involved in oncogenesis were collected from CancerGene database which contains 3164 proteins which are thoroughly curated with information from key publicly available database [60]. The human cancer gene network (HCGN) was constructed by mapping Human PPI obtained from HIPPIE against cancer genes of CancerGene database which then

consisted of 1,694 proteins with 8,672 interactions. After removal of orphan nodes from the HCGN, the giant component culminates with 8,668 interactions among 1,686 proteins.

Cervical cancer gene dataset

The cervical cancer gene dataset was obtained from the cervical cancer gene database that catalogs information of genes associated with cervical cancer. CCDB (Cervical Cancer Gene Database) consisting of 538 genes is a specialized, manually curated database that contains information of all experimentally determined cancer genes that are involved in human cervical carcinogenesis [61]. The genes that were found to be common in both the cervical cancer and human PPI datasets were enumerated as 176 and were considered for our further analysis.

Topological properties of HCGN

The Human cancer gene network was analyzed for their topological properties such as degree, efficiency, diameter and average clustering coefficient. The importance of a node in the network structure is quantified in terms of centrality measure. Different centrality measures focus on different importance concepts and are categorized in to 6 types based on different concepts of ranking such as neighborhood, distance, shortest path, current flow, feedback and vitality in case of biological networks. Here we have calculated six different graph centrality measures such as vulnerability, closeness, centroid values, shortest-path betweenness, current-flow betweenness, and Eigen vector using the tool CentiBin and are defined as follows [43,62].

*Vulnerability***:** This centrality measure is calculated from the change in efficiency when a particular node is knocked out from the network. The efficiency of the network is measured from the inverse sum of the distance matrix. The shortest path of communication between any two nodes is termed as distance matrix. The efficiency of the network is defined as

$$E(G) = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_{i,j}} \quad (1)$$

Here G is the graph, $d_{i,j}$ is the shortest path between node i and j, and N(N-1) is the normalization constant. The vulnerability of the node v is calculated from

$$C_v(v) = E(G) - E(G - v_i) \quad (2)$$

Closeness centrality: Closeness C_c (v) is defined as the reciprocal of the total distance from a node v, to all other nodes. It is given by,

$$C_c(v) = \frac{1}{\Sigma_{u \epsilon v} dist(u,v)} \quad (3)$$

Centroid values: The centroid value is the most complex node centrality index and is computed by focusing the calculus on couples of nodes (v,w) and systematically counting the nodes that are closer (in term of shortest path) to v or to w. A node v with the highest centroid value is the node with the highest number of neighbors separated by the shortest path to v.

$$C_{cen}(v) = min\{f(v,w): V\{v\}\} \quad (4)$$

Where f (v, w) = γ v (w) – γ w (v) and γ v (w) denotes the number of vertices that are closer to v than to w.

Shortest path betweenness centrality: Shortest path betweenness represents the contribution of a node v, towards communication between all nodes pairs. It is defined as,

$$C_B(v) = \Sigma_{s \neq t \neq v \neq V} \frac{\rho_{st}(v)}{\rho_{st}} \quad (5)$$

Current flow betweenness centrality: Current flow betweenness of a node v is the average of the current flow over all source-target pairs.

$$C_{CB}(v) = \frac{\Sigma_{s \neq t \in V} I_v{}^{(st)}}{\frac{1}{2}n(n-1)} \quad (6)$$

Eigen vector centrality: scores the relative importance of all nodes in the network by weighting connections to highly important nodes more than connections to nodes of low importance. It can be calculated by

$$\lambda C_{IV} = AC_{IV} \quad (7)$$

Where, C_{IV} denotes the Eigen vector and λ denotes the Eigen value.

Correlation analysis of centrality measures

The six different centrality measures were calculated for each and every node in the interactome and ranked based on their scores. Pair wise correlation between the various centrality measures was obtained through Spearman's rank correlation coefficient ρ which is defined as

$$\rho = 1 - \frac{6\Sigma {d_i}^2}{n(n^2-1)} \quad (8)$$

Here, the difference d_i represents the difference in the ranks of each observation on the two variables which here represents the centrality scores. Also for each centrality measures the top 50 ranking gene set is collected and the dataset is pooled into a single list which was further utilized in our study for prioritizing the candidate genes.

Gene ontology enrichment analysis

The gene ontology enrichment analysis was performed with the help of GOrilla which manipulate the flexible threshold statistical approach to determine the GO terms that are significantly enriched at the top of ranked gene list [63].The significantly enriched GO terms were obtained using two ranked gene list mode with the list of known cervical cancer genes as a

target and HCGN genes as a background. The significantly enriched GO terms for Biological process (BP), Molecular function (MF) and Cellular components (CC) were achieved with a p value <0.001 and the number of genes that are associated with a specific GO terms (3≤B≤50). The range for number of genes associated with a specific GO term is chosen in such a way that it scrutinizes only the GO terms which were annotated for at least three but not more than 50 genes. The value of B is considered an important aspect for the gene enrichment analysis because high value of B increases noise by populating the enrichment results with non-specific GO terms whereas small values of B reduce the signal by rejecting specific GO terms [47, 64].Thus p-value threshold (<0.001) and the B- value (3≤B≤50) were chosen in such a way to maximize the signal (specific GO terms) and reduce the noise (non-specific GO terms) in the enrichment analysis. The genes associated with the significantly enriched GO terms of BP, MF and CC were found to be analogous to the corresponding GO terms of HCGN genes and henceforth acknowledged as the genes with specific disease ontologies.

Prediction of candidate genes for cervical cancer

To predict the candidates of cervical cancer three gene sets A, B and C were analyzed, where the Set A consists of genes obtained from amalgamating the top 50 ranking genes of each of the six different centrality measures. The set B and C were composed of the known cervical cancer genes and genes with specific disease ontologies respectively. The top50 ranking pooled gene list obtained from the six different centrality measures is related with the known cervical cancer genes which in turn it is correlated with the significant disease ontology genes retrieved from gene enrichment analysis. The genes that are mutual to top50 ranking genes and significant disease ontology but not generic to cervical cancer are depicted as cervical cancer candidates.

Results

Human cancer gene network and its topological properties

With the above theoretical approach, the HCGN was constructed in such a way that the cancer genes contributing to carcinogenesis accumulation was to form a sub network within the HIPPIE dataset. The HCGN is constructed by mapping HIPPIE dataset of 70,401 interactions among 11,468 proteins against the 3,164 catalogued cancer genes. The network has 8,672 edges among 1,694 nodes. The orphan nodes of HCGN are removed and the core network is encompassed with 8,668 interactions between 1,686 proteins.

To define the interaction network its topological parameters like degree, diameter, correlation, efficiency, etc., have a pivotal role in enhancing them. Precisely, for the HCGN the average degree, the diameter, assortative correlation and global efficiency were found to be 10.28, 9, 0.44, and 0.31 respectively. The degree value, as expected follows a power law distribution with an exponent of 2.23 (**Figure-2**) and the average clustering coefficient for the HCGN network was found to be 0.1698. The interaction network concerned at the molecular level is considered as scale-free in nature since it has been marked by the presence of hubs and also is evident from the scaling exponent.

Graph centralities of HCGN

In our study we have utilized six different centrality measures to ascertain the potentiality of individual proteins in HCGN. After consolidation of top 50 ranking gene sets of the six different centrality measures 92 (set A) genes were obtained which are nothing but the representative generic cancer genes that can be used to predict the candidate genes for the cervical cancer. The pair wise correlation coefficients of the six centrality measures depicted for the HCGN elucidate

that they all are positively correlated as represented in **Table-1**. It also elucidates that the ranking of the nodes differs based on the formulary of each centrality measure.

GO enrichment analysis of the top 50 ranking genes

The gene ontology enrichment analysis was performed for the top 50 ranking genes in order to justify the role of centrality measures in identifying the significant cervical cancer related gene ontological terms. The genes were enriched using GOrilla software by taking 92 top 50 high ranking genes (set A) as a target and HCGN genes as a background which then yielded 290 disease specific ontology genes (set C).The GO terms enriched for all the three domains Biological Process, Molecular function and Cellular component were obtained and were enumerated as 92. The cellular component contains only six ontology terms whereas the molecular function contains 21. Comparatively the biological process preponderate the cellular components and molecular function's gene ontologies. Among the cellular component IkappaB kinase complex (GO: 0008385) and CD40 receptor complex (GO:0035631) were dominant over the rest with the enrichment values of 13.66and 10.41 respectively. But comparing the biological process and molecular function the cellular component remained suppressed in case of top50 ranking genes. While considering the 21 GO terms of molecular function the binding factor ontological terms dominated the activity related GO terms in terms of their count. Interestingly nitric-oxide synthase regulator activity and IkappaB kinase activity have a high enrichment value of 18.21. A list of 71 significantly enriched GO terms was associated with the biological process for pooled top50 ranking HCGN genes where activity, reputational processes and the signaling pathways were numerous. Among the biological processes, B cell lineage commitment (GO: 0002326) and primary miRNA processing (GO: 0031053) were observed to have high enrichment score of 18.21. Most of the ontological terms prevailing in our list such as SH2

domain binding (GO:0042169) ,core promoter binding (GO:0001047), primary miRNA processing (GO:0031053),G1 DNA damage checkpoint (GO:0044783), mitotic G1 DNA damage checkpoint(GO:0031571), protein import into nucleus, positive regulation of apoptotic signaling pathway (GO:2001235), nuclear transport(GO:0051169) has been associated with various cell cycle process, growth signaling pathways which are capable of altering the intracellular mechanisms that are capable of causing cancer. The detailed list of GO terms for the HCGN genes is given in **Supplementary data S1**.

GO enrichment analysis of known cervical cancer genes

The GO enrichment for the known cervical cancer genes is performed through GOrilla by claiming the known cervical cancer genes of 176 (set B) as the target and HCGN gene set of 1,686 as a source. The analysis is performed with a p-value < 0.001. It resulted in 102 GO terms from all the three ontologies BP, MF and CC with the B-value ($3 \leq B \leq 50$). The biological process was dominant over the molecular function and the cellular component. The cellular component contained only two GO terms proteinaceous extracellular matrix (GO: 0005578) with the enrichment value of 3.58 and extracellular matrix (GO: 0031012) with the value of 3.56. The molecular function contained 11 GO terms among which the different binding terms were dominant over the activity. The biological process contained 89 GO terms among which the regulatory terms were found to be dominant over the rest. We have found out that our list enriched ontological terms for known cervical cancer genes were confederated with the cycle, apoptotic and regulation process. Some of them were as follows negative regulation of cell morphogenesis involved in differentiation (GO: 0010771), positive regulation of intracellular transport (GO: 0032388), positive regulation of chemokine production (GO: 0032722), cellular response to acid (GO:0071229),negative regulation of cell growth (GO:0030308), positive

regulation of cell adhesion (GO:0045785), cell-cell adhesion (GO:0016337). The detailed list of enriched ontological terms for known cervical cancer genes is provided in the **Supplementary data S2**.

Genes with high-network scores and significant GO terms as predicted candidate genes

Towards predicting candidate genes for cervical cancer which are of therapeutic value, we rationally correlated the three major set of genes. The set A is from the pooled top 50 ranked genes list whereas the set B is the cluster of known cervical cancer genes and the set C is a set of genes with significantly enriched disease ontologies. Altogether, set A contains 92 genes among which set A and set B share 24 genes in common while set B contains 176 cervical cancer genes. The set C contained 290 genes obtained from 102 GO terms where the set B and set C shared 79 genes in common. 24 genes participated among set A and set B while set A and set C shared 29 genes in common. Those genes that are common to top50 ranking gene list and genes with significant disease ontologies but were not amid the known cervical cancer genes were identified to be the candidate genes for cervical cancer. The genes of the three sets are logically juxtaposed which represents the strategy employed for predicting the candidate genes for cervical cancer and the same is depicted in the Venn diagram **Figure-3**. The candidate genes for cervix related carcinogenesis is estimated to be 15 which are unique and neither found as common in any of the sets. Also we have carried out the same analysis with the three centrality measures degree, closeness and vulnerability as specified in Casaba and coworkers approach of candidate gene prediction. The genes that were predicted to serves as candidates using both the approaches were compared and found out that of the 15 candidate genes predicted through the six different centrality measures were in common with the 13 candidate genes obtained using 3 centralities. Our approach was able to predict two genes HIF1A and RET in addition as candidate genes for

cervical cancer and thus elucidates the importance of usage different centrality measures in candidate gene prediction.

The predicted candidates of cervical cancer

The 15 potential protein targets identified for cervical cancer were explored to find the disease relevance for their distinctive role in cervical cancer and was discovered that they are somehow significant to the carcinogenic advance in a cell. Literatures cram for the identified candidates of the cervical cancer helps in analyzing how important the predicted disease gene is. The list of genes prioritized for cervical carcinogenesis along with their description is given in **Table-2**. Among the predicted 15 novel candidate genes, the gene EP300 commonly known as p300 is involved in pathways of cancer and has a foremost role in the process of cell proliferation and differentiation. EP300 is concerned with few key functions as inhibition of apoptosis, proliferation and accumulation of mutation. The JUN gene is a putative transformation gene which takes part in the transformation pathways of cancer. The protein encoded by the SMAD3 gene functions as a transcriptional modulator that regulate the carcinogenic onset where as the gene CAV1 was found to be a tumor suppressor gene candidate.

The phosphoprotein PML gene functions as a transcription factor and a tumor suppressor. This gene regulates the p53 response to oncogenic signals that have an escort role in cervical cancer through p53 signaling pathway. Amplification of ERBB3 gene or overexpression of this protein has been reported in numerous cancers where the heterodimerization of it leads to activation of pathways which in turn leads to cell proliferation and differentiation. The proto-oncogene SRC has an extensive role in regulation of embryonic development and cell growth. Any mutations in this gene could be involved in the malignant progression of cancer. With the above summary, it

can be relished that the predicted candidates have a significant role in carcinogenesis and special attention could be drawn towards identifying potential drug targets for the cervical cancer.

Discussion

In recent years, protein interaction networks are primarily used in targeting genes responsible for a disease. Towards identifying candidate protein target essential for cervix related carcinogenesis, we used an integrative network and gene ontological approach. The network properties provides a system perspective of complex molecular mechanism and helps to identify the functional elements while the gene enrichment analysis helps to identify the ontological features of a gene set. In general the disease gene prioritization is a difficult task through wet lab experiments which perpetuate for generations. But the computational method for predicting disease gene is achieved through various methods where protein interaction network is in vicinity towards researchers. Network analysis is a potent approach in understanding the disease phenotype and probing for therapeutic targets [55]. Also the functional importance of the protein can be distinguished from the network through the centrality measures.

Earlier, Csaba Ortutay when predicting candidate genes used only three centrality measures along with GO terms. But in our analysis we have used six different centralities which were the efficient tools for network analysis for predicting cervical cancer candidates. The edge gained by using six different centrality measures is that almost all the biologically prominent genes were obtained in either of the top ranking genes of each centrality measure which are in confirmative with the gene enrichment analysis. The six different centralities were calculated for 1,686 proteins in the interactome and the scores of these measures show a strong correlation and it is also used to quantify the importance of protein in the interactome. It is worth mentioning, we

have used CentiBiN tool to calculate the centrality measures for the Human cancer gene network. This tool has 17 centrality measures out of which we were able to calculate only 13 centrality measures. The vulnerability centrality measure which is not available in CentiBiN tool was calculated through MATLAB programming. From these 14 centrality measures, we have considered the measures whose pair-wise correlation coefficient with other centrality measures is less than 1. Although the ranking of nodes are different in the calculated centrality measures but in our study, we are considering and pooling the top 50 ranking genes, the ranking position is not important. In this study only six centrality measures such as vulnerability, closeness, centroid values, shortest path betweenness, current flow betweenness and Eigen vector are sufficient. It is important to note that the combination of centrality measures may vary with the percentage of top ranking genes and the network we study.

Gene ontology enrichment analysis provides means of identification of significantly overrepresented GO terms which could be effectively used to get biological insight from a given set of genes [64]. The biological relevance of a protein can be extracted from the Gene ontology terms which provide information through the BP, MF and CC terminologies. In our work we have predicted the candidate genes for cervical cancer employing the approach of Csaba Ortutay but with more number of centrality measures. 15 novel cervical cancer candidate genes were prioritized by logically juxtaposing the Set A obtained from the result of genes pooled through top 50 ranking genes of the six centrality measures, the set B with known cervical cancer genes and the set C containing the genes with significant disease ontologies. Validation of the predicted candidates is indispensable to conclude them as a potential target for a disease state. The predicted cervical cancer genes were analysed through literature survey to prove them that they

can act as a targets for cervical carcinogenesis. The annotation of the predicted genes for vindication of their disease relevance is as follows.

The gene ARRB1 also known as β-arrestin 1was found not to have direct implication for cervical carcinogenesis but it is overexpressed in gastric cardiac adenocarcinoma as is evident from the Wang et al work [65]. A recent study states that, the human papillomavirus that cause the cervical cancer has been linked with an increased risk of cardiovascular diseases. ARRB2 member of beta-arrestin protein family was shown to inhibit beta-adrenergic receptor function. Recently many studies have revealed that this gene may act as an adapter for scaffolding many intracellular signalling networks that may lead to cancerous conditions.Understanding the role of these β-arrestins in carcinogenesis is highly complicate because of their complex biological and regulation events [66]. A better knowledge regarding the prognosis and oncogenic potential of β-arrestins encumbrances the identification of potential candidate genes for various tumours including carcinoma of cervix. Whereas the CAV1, Ceaveolin-1 gene, the main component of caveolae plasma membrane is found in most of the cell types that can be regarded as candidate for tumor suppressor and it is over expressed in terminally differentiated cells. CAV1 contributes to tumorigenesis of cervical cancer due its down regulation in cells transformed by HPV infection [67].

The CFTR gene is primarily involved in the transport of chloride ions. Peng along with his co-workers from their studies on cervical cancer suggested that CFTR may act as potential therapeutic target for cervical cancer because of its higher expression levels [68]. EP300 which is also designated as p300 gene plays a crucial role in cell differentiation and mutational events

and any abnormalities in this gene contributes to carcinogenesis. Stina and group evaluated the role expression of p300 in the outcome of cervical cancer where the immunohistochemistry study revealed that the transcription factor p300, was up regulated in cervical intra Epithelial neoplasia [69] .The gene ERBB3, (V-erb-b2 avian erythroblastic leukaemia viral oncogene homolog3 gene) encodes a member of epidermal growth factor receptor family of receptor tyrosine kinases. Being the third member of the ErBB proto oncogene family, c-erbB-3 (ErBB3) is toughly expressed and amplified in numerous cancers. The immunohistochemical study carried out by Hunt and his co-workers have put forward that c-erBB-3 is widely expressed in cervical carcinomas [70]. The transcriptional factor HIF1A, Hypoxia-inducible factor 1α (HIF-1α) contributes tumor growth and progression through promotion of neoangiogenesis and regulation of the genes involved in response to hypoxia. Birner and his co workers in their work have proven that HIF-1α expression is a strong independent prognostic marker in early stage cervical cancer [71].

The Insulin receptor gene, INSR has important roles in cancer. As Serrano and his co-workers report, the receptor expression was diverse that the tyrosine phosphorylation of them is correlated with high expression level [72]. However they show no effect on proliferation, migration or invasion of the cell line. The genes JAK2 (Janus kinase2) and JUN (jun proto-oncogene) are involved in various processes such as cell growth development and differentiation was found to have and altered gene expression in cervical cancer as is evident from Carlos et al work. JAK2 the protein tyrosine kinase involved in JAK-STAT pathway was found to be down regulated and JUN of focal adhesion pathway was overexpressed with the ratios of -2.9 and 4.8 respectively[73].

The gene LYN (V-Yes-1 Yamguchi Sarcoma viral related oncogene) plays an important role in the regulation of innate and adaptive immune responses. LYN signalling may play a vital role in survival and proliferation of some types of cancer cells. The patent of Iftner et al has produced a list of diagnostic markers for determining the genetic and environmental factors for cervical carcinogenesis. LYN was found to be one of the diagnostic markers among the list of genes that contributes to cervical cancer due to HPV infection [74]. PML, the protein encoded by Promyelocytic leukaemia gene is a member of tripartite motif family and regulates the P53 response to oncogenic signals. PML reinforces carcinogenesis by exhibiting a synergetic action with the HPV infection, the main convict causing cervical cancer. This is evident from the observations drawn by Neha Singh and group suggesting that down regulation of PML gene coupled with HPV infection contributes to cervical carcinogenesis [75].

The RET proto oncogene is found to be involved in the tumourigenesis of thyroid carcinoma. As described by Vamsy and coworkers that the squamous cell carcinoma of the cervix is functionally cured by the rare phenomena of metastatic thyroid carcinoma that carries a RET gene as its major contributor [76]. This lead us to a conclusion that RET proto oncogenes can indirectly donate to the cervical cancer. We believe that it has role in cervical carcinoma and this may be experimentally tested. The SMAD3 gene demonstrates that disruption of TGF-beta/Smad signaling pathway exists in human cervical cancer and over expression of it may contribute malignant progression of human cervical tumours [77]. The gene SRC (v-src avaian sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog), also known as C-SRC is a proto oncogene. Over expression of SRC has been associated with enhanced cancer cell growth [78]. Over

expression of phosphorylated SRC has been found in the cervical cancer cell lines and clinical cervical cancer tissue [79]. Recently Teng and his co-workers in their study have ascertained that SRC signalling play an essential role in cervical cancer progression [80].

The predicted genes were searched against the DrugMap Central and Therapeutic target database to identify the available drugs had either formerly served as an objective for cervical cancer and to analyse its metabolic pathway. The predicted candidate gene JUN, JAK2, INSR, SMAD3, ERBB3, SRC were all searched against TTD [81]. All these genes have been identified as either clinical trial or research or successful targets for major diseases like cancer, diabetes, vascular disease. These genes could also be a potential target for the cervical cancer. The JUN gene is involved in pathways related to cancer, renal cell carcinoma and diverse signalling pathways. The genes JAK2, INSR, ERBB3, and SRC reported to be the candidates of the cervical cancer and validated through TTD database is involved in cancer pathway and also they have an extensive role in signalling pathways. This information is summarized in **Supplementary data S3**.

The genes such as CFTR, and LYN have been predicted to be the targets for drug through DMC [82]. The cystic fibrosis Trans membrane conductance regulator (CFTR) gene is involved in pathways such as ABC transporters, bile secretion, pancreatic secretion, gastric acid secretion. The LYN gene is involved in ATP binding and is seen in chemokine signalling pathway, B cell receptor signalling pathway, Fc epsilon RI signalling pathway. These targets could be analysed for the cervical cancers too to identify the drugs for the cervix related carcinogenesis.

The genes identified as potential has already been either a clinical or research or successful target for a number of diseases primarily cancer and this in turn could also be analysed for the cervix related oncogenesis. The protein-protein interaction network with the cancer genes available has paved the way for identifying the candidate genes for cervix related carcinogenesis through network properties and gene ontologies. In our work we have used six different graph centrality measures rather than three as used by Csaba Ortutay in his work. Various centrality measures ranks the nodes based on different concepts such as neighbourhood, distance shortest path etc and thereby abstracts the potential candidate genes. This is evident from our work that the genes LYN, ERRB3 scored among the top ranking genes for only of Eigenvector centrality were proven to be the potential genes for cervical carcinogenesis. The presence of the candidate genes along in the respective centrality measure are identified and are presented as heatmap in **Figure-4**. Also the predicted 15 candidate genes and their interacting partners are given in the **Supplementary data S4**.

Apart from literature survey, we have also tried to find out the biological relevance of the predicted candidate genes by analyzing the pathways of cervical cancer caused due to viral carcinogenesis. The cervical cancer pathway (hsa05203) collected from the KEGG disease database projects that the gene SRC is involved in the MAPK signalling pathway of the cervical cancer caused by Hepatitis B virus. Similarly the gene LYN plays a crucial role in cell receptor signalling pathway of cervical cancer due to Epstein Barr virus by inhibition of apoptosis, where as the gene EP300 inhibits P53 and thereby inhibits the apoptosis, proliferation and accumulation of mutations in P53 signaling pathway of cancer due to Human papilloma virus. Also we have tried to validate the biological relevance of the predicted candidate genes using protein protein

interactions utilizing STRING data base [83]. We have submitted the list of the candidate genes individually and together to the database and compared the results with the interactions of our approach. Interestingly we have found out that the results matched with our interaction pairs and the same are highlighted in the **Supplementary data S4**. The results of the KEGG pathway enrichment analysis done in STRING data base confirms the relevance of interaction pairs to the disease, which indirectly depicts the contribution of predicted candidate genes for cervical carcinogenesis.

To strengthen our findings, the performance of the method was carried out using the leave-one-out statistical test. The analysis was repeated for 176 times (the number of known cervical cancer genes) by leaving one known cervical cancer gene out at time. We found that all the 14 known cervical cancer genes which are also present in SET A and C, were identified as disease related. The statistical test resulted 88% performance for the genes that are present in SET A, B and C (see **Figure-3**). Thus the predicted 15 candidate genes with high network score and significant disease ontologies might have relevance to cervix related cancer. The performance test was implemented through MATLAB programming.

Thus the GO terms coupled with usage of six different centrality measures contributed in successful prioritization of 15 novel cervical cancer candidate genes. Among the 15 predicted genes, the genes EP300, SRC and SMAD3 were present in all the six centrality measures. Interestingly, all these three genes were among top 15 in their ranking with in all the six centrality measures. Also from the literature survey, these three genes were proven to more successful in causing cervical cancer which implies that they can act as better candidates compared to the rest 12 genes for cervical carcinogenesis. The predicted genes which were

proved for their role in cervical carcinogenesis could be searched for the drugs and may serves as a potential drug targets for cancer of cervix. Thus through our analysis we have procured 15 novel candidate genes for cervical carcinogenesis which might facilitate the identification of diagnosis biomarkers and development of drug targets and thereby boosts up the cervical carcinoma research.

Acknowledgements:

The authors VH, TK, SBR and PM would like to thank Dept. of Science and Technology, Government of India, for their financial support (DST-CMS GoI Project No. SR/S4/MS: 516/07 Dated 21.04.2008).

References

1. M. Sharma, and P.K. Majumdar, *Indian J Occup Environ Med*. 2009, **13**,109-112.
2. B. Vogelstein, and K.W. Kinzler, *Nat. Med.* 2005, **10**, 789–799.
3. D. Hanahan, and R.A. Weinberg, *Cell*, 2011, **144**, 646-674.
4. D. Hahanan, R.A. Weinberg, and S. Francisco. *Cell*, 2000, **100**, 57-70.
5. J. Sherris, C. Herdman, and C. Elias. *West J Med.* 2001, **175**, 231–233.
6. T. Rajkumar, K. Sabitha, N. Vijayalakshmi, S. Shirley, M.V. Bose, G. Gopal, and G. Selvaluxmy, *BMC Cancer* 2011, **11**,80.
7. M.G. Kim, F.A. Flomerfelt, K.N. Lee, C. Chen, and R.H. Schwartz, *J. Immunol.* 2000, **164**, 3185-3192.
8. T. Strachan, and A.P. Read*, Human Molecular genetics*, New York: Wiley- Liss; 1999.
9. T. Yoshida, and K. Yoshimura, *Proc. Jpn. Acad.* 2003, **79**, 34–50.
10. K. Schubert, H. von Bonnsdorf, M. Burke, I. Ahlert, S. Braun, R. Berner, K.A. Deichmann, and A. Heinzmann, *Dis. Markers* 2006, **22**, 127–132.
11. T. Miyata, *Hypertens Res.* 2008, **31**,173–174.

12. S. Kohler, S. Bauer, D. Horn, and P.N. Robinson. Am J Hum Genet. 2008, 82, 949–958.
13. X. Wu, R. Jiang, M.Q. Zhang, and S Li. Mol Syst Biol. 2008, 4, 189.
14. H.K. Tabor, J.R. Neil, and M.M. Richard. *Nat Rev Genet* 2002, **3**, 391-397.
15. M. Zhu, and S. Zhao, *Int J Biol Sci.* 2007, **3**, 420–427.
16. M.J. Zhu, X. Li, and S.H. Zhao, *Methods Mol Biol.* 2010, **653**, 105-129.
17. K.M. Giacomini, C.M. Brett, R.B. Altman, N.L. Benowitz, M.E. Dolan, D.A. Flockhart, J.A. Johnson, D.F. Hayes, T. Klein, R.M. Krauss, D.L. Kroetz, H.L. McLeod, A.T. Nguyen, M.J. Ratain, M.V. Relling, V. Reus, D.M. Roden, C.A. Schaefer, A.R. Shuldiner, T. Skaar, K. Tantisira, R.F. Tyndale, L. Wang, R.M. Weinshilboum, S.T. Weiss, and I. Zineh, *Clin Pharmacol Ther* 2007, **81**, 328–345.
18. R.D. Hawkins, G.C. Hon, and B. Ren, *Nat Rev Genet* 2010, **11**, 476–486.
19. Y.A. Kim, S. Wuchty, and T.M. Przytycka, *PLoS Comput Biol.* 2011, **7**, e1001095.
20. C. Perez-Iratxet, P. Bork, and M.A. Andrade, *Nat Genet* 2002, **31**, 316-319.
21. M. Pellegrini-Calace, and A. Tramontano, *Bioinformatics.* 2006, **22**, 775-778.
22. E.A. Adie, R.R. Adams, K.L. Evans, D.J. Porteous, and B.S. Pickard, *BMC Bioinformatics* 2005, **6**, 55.
23. T.D. Bie, L.C. Tranchevent, L.M.M van Oeffelen, and Y. Moreau, *Bioinformatics* 2007, **23**, i125-i132.
24. J. Freudenberg, and P. Propping. *Bioinformatics* 2002; **18**, S110-S115.
25. N. Sugaya, K. Ikeda, T. Tashiro, S. Takeda, J. Otomo, Y. Ishida, A. Shiratori, A. Toyod, H. Noguchi, T. Takeda, S. Kuhara, Y. Sakaki, and T. Iwayanagi, *BMC Pharmacol.* 2007, **7**, 10
26. L. Franke, H. van Bakel, L. Fokkens, E.D. de Jong, M. Egmont-Petersen, and C. Wijmenga, *Am J Hum Genet.* 2006, **78**, 1011-1025.
27. S. Rossi, D. Masotti, C. Nardini, E. Bonora, G. Romeo, E. Macii, L. Benini, and S. Volinia, *Nucleic Acids Res* 2006, **34**, W285–W292.
28. R.A. George, J.Y. Liu, L.L. Feng, R.J. Bryson-Richardson, D. Fatkin, and M.A. Wouters, *Nucleic Acids Res* 2006, **34**, e130.
29. A.L. Yonan, A.A. Palmer, K.C. Smith, I. Feldman, H.K. Lee, J.M. Yonan, S.G. Fischer, P. Pavlidis, and T.C. Gilliam, *Genes Brain Behav.* 2003, **2**, 303-320.
30. E.A. Adie, R.R. Adams, K.L. Evans, D.J. Porteous, and B.S. Pickard, *Bioinformatics* 2006, **22**, 773-774.

31. F.S. Turner, D.R. Clutterbuck, and C.A. Semple, *Genome Biol* 2003, **4**, R75.
32. C. Perez-Iratxeta, M. Wjst, P. Bork, and M.A. Andrade, *BMC Genet*. 2005, **6**, 45.
33. P Zhang, J. Zhang, H. Sheng, J.J. Russo, B. Osborne, and K. Buetow, *BMC Bioinformatics*. 2006, **7**, 135.
34. S. Aerts, D. Lambrechts, S. Maity, P. van Loo, B. Coessens, F. De Smet, L.C. Tranchevent, B. De Moor, P. Marynen, B. Hassan, P. Carmeliet, and Y. Moreau, *Nat Biotechnol.* 2006, **24**, 537-544.
35. L.S. Collier, and D.A. Largaespada, *Curr Opin Genet Dev* 2006, **16**, 23–29.
36. J.B. Bard, and S.Y. Rhee, *Nat Rev Genet* 2004, **5**, 213-222.
37. J. Chen, H. Xu, B.J. Aronow, and A.G. Jegga, *BMC Bioinformatics* 2007, **8**, 392.
38. A.-L. Barabási, N. Gulbahce, and J. Loscalzo, *Nat. Rev. Genet.* 2011, **12**, 56-68.
39. E. Wang, J. Zou, N. Zaman, L.K. Beitel, M. Trifiro, and M. Paliouras, *Semin Cancer Biol* 2013, **23**, 279-285.
40. P. Holme, M. Huss, H. Jeong, *Bioinformatics* 2003, **19**, 532–538.
41. S. Wuchty, and P.F. Stadler, *J Theor Biol* 2003, **223**, 45-53.
42. P. Manimaran, S.R. Hegde, and S.C. Mande, *Mol BioSyst* 2009, **5**, 1936-1942.
43. A. Zhang, *Protein Interaction Networks: Computational Analysis.* Cambridge University Press, 2009.
44. T. Kranthi, S.B. Rao, and P. Manimaran, *Mol BioSyst* 2013, **9**, 2163-2167.
45. G. Caldarelli, *Scale-Free Networks: Complex webs in nature and technology,* Oxford University Press; 2007.
46. P. Manimaran, S.R. Hedge, and S.C. Mande, *PLoS Comput Biol* 2008, **4**, e1000237.
47. U. Stelzl, U. Worm, M. Lalowski, C. Haenig, F.H. Brembeck, H. Goehler, M. Stroedicke, M. Zenkner, A. Schoenherr, S. Koeppen, J. Timm, S. Mintzlaff, C. Abraham, N. Bock, S. Kietzmann, A. Goedde, E. Toksöz, A. Droege, S. Krobitsch, B. Korn, W. Birchmeier, H. Lehrach, and E.E. Wanker, *Cell* 2005, **122**, 957-968.
48. J. Chen, B.J. Aronow, and A.G. Jegga, *BMC Bioinformatics* 2009, **10**, 73.
49. G. Wu, X. Feng, and L. Stein, *Genome Biol* 2010, **11**, R53.
50. G. Kar, A. Gursoy, and O. Keskin, *PLoS Comput Biol.* 2009, **5**, e1000601.
51. M.G. Kann, *Brief Bioinform* 2007, 8, 333–346.

52. K.I. Goh, M.E. Cusick, D. Valle, B. Childs, M. Vidal, and A.L. Barabási, *Proc Natl Acad Sci USA*. 2007; 104:8685-90.
53. P.F. Jonsson, and P.A. Bates, *Bioinformatics* 2006, **22**, 2291–2297.
54. S. Wachi, K. Yoneda, and R. Wu, *Bioinformatics* 2005, **21**, 4205–4208.
55. C. Ortutay, and M. Vihinen, *Nucleic Acids Res* 2009, **37**, 622-628.
56. M.H. Schaefer, J.F. Fontaine, A. Vinayagam, P. Porras, E.E. Wanker, M.A. Andrade-Navarro, *PLoS One* 2012, **7**, e31826.
57. T.S.K. Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D.S. Somanathan, A. Sebastian, S. Rani, S. Ray, C.J.H. Kishore, S. Kanth, M. Ahmed, M.K. Kashyap, R. Mohmood, Y.L. Ramachandra, V. Krishna, B.A. Rahiman, S. Mohan, P. Ranganathan, S. Ramabadran, R. Chaerkady, and A. Pandey, *Nucleic Acids Res* 2009, **37**, D767-D772.
58. A. Chatr-aryamontri, A. Ceol, L.M. Palazzi, G. Nardelli, M.V. Schneider, L. Castagnoli, and G. Cesareni, *Nucleic Acids Res* 2007, **35**, D572-4.
59. I. Xenarios, D.W. Rice, L. Salwinski, M.K. Baron, E.M. Marcotte, and D. Eisenberg, *Nucleic Acids Res.* 2000, **28**, 289-291.
60. M.E. Higgins, M. Claremont, J.E. Major, C. Sander, and A.E. Lash, *Nucleic Acids Res.* 2007, **35**, D721-D726.
61. S.M. Agarwal, D. Raghav, H. Singh, and G.P.S. Raghava, *Nucleic Acids Res* 2011, **39**, D975-D979.
62. H.B. Junker, D. Koschützki, and F. Schreiber, *BMC Bioinformatics* 2006, **7**, 219
63. E. Eden, R. Navon, I. Steinfeld, D. Lipson, and Z. Yakhini, *BMC Bioinformatics* 2009, **10**, 48.
64. S. Vashisht, and G. Bagler, *PLoS One* 2012, **7**, e49401.
65. L.G. Wang, B.H. Su, and J.J. Du, *Asian Pac J Cancer Prev* 2012, **13**, 5671-5675.
66. S. Hu, D. Wang, J. Wu, J. Jin, W. Wei, and W. Sun, *Mol Biol Rep*. 2013, **40**, 1065-1071.
67. T.F. Chan, T.H. Su, K.T. Yeh, J.Y. Chang, T.H. Lin, J.C. Chen, S.S. Yuang, and J.G. Chang, *Int J Oncol* 2003, **23**, 599-604.
68. X. Peng, Z. Wu, L. Yu, J. Li, W. Xu, H.C. Chan, Y. Zhang, and L. Hu, *Gynecol Oncol*. 2012, **125**, 470-476.

69. S. Syrjänen, P. Naud, L. Sarian, S. Derchain, C. Roteli-Martins, A. Longatto-Filho, S. Tatti, M. Branca, M. Erzen, L. Serpa-Hammes, J. Matos, F. Arlindo, M. Sakamoto-Maeda, S. Costa, and K. Syrjänen, *Int J Gynecol Pathol* 2010, **29**, 135-145.
70. C.R. Hunt, R.J. Hale, C. Armstrong, T. Rajkumar, W.J. Gullick, C.H. Buckley, *Int J Gynecol Cancer.* 1995, **5**, 282-285.
71. Birner, P., Schindl, M., Obermair, A., Plank, C., Breitenecker, G., & Oberhuber, G. *Cancer Research*, 2000, **60**, 4693-4696.
72. M.L. Serrano, M. Sánchez-Gómez, M.M. Bravo, S. Yakar, and D. LeRoith, *Horm Metab Res* 2008, 40, 661-667.
73. C. Pérez-Plasencia, G. Vázquez-Ortiz, R. López-Romero, P. Piña-Sanchez, J. Moreno, and M. Salcedo. *Infect Agent Cancer* 2007, **2**, 16.
74. T. Iftner, F. Stubenrauch, A. Manawapat, and S.K. Kjaer, *Patent Appln:US 13/294*, 905 (2012).
75. N. Singh, R.C. Sobti, V. Suri, R. Nijhawan, S. Sharma, B.C. Das, M. Bharadwaj, S. Hussain, *Gynecol Oncol.* 2013, **128**, 420-426.
76. M. Vamsy, P.S. Dattatreya, L.Y. Sarma, M. Dayal, N. Janardhan, and V.V.S.P. Rao, *Indian J Nucl Med.* 2013; **28**, 112–114.
77. K.D. Ki, S.Y. Tong, C.Y. Huh, J.M. Lee, S.K. Lee, and S.G. Chi, *J Gynecol Oncol* 2009; **20**, 117-121.
78. S. Karmakar, E.A. Foster, C.L. Smith, *Endocrinology* 2009, **150**, 1588–1596.
79. S.J. Cheng, S.H. Kok, J.J. Lee, M.Y.P. Kuo, S.L. Cheng, Y.L. Huang, H.M. Chen, H.H. Chang, and C.P. Chiang, *Head Neck*. 2012, **34**, 1340–1345.
80. T. Hou, J. Xiao, H. Zhang, H. Gu, Y. Feng, and J. Li, *Int J Clin Exp Pathol* 2013, **6**, 1121-1127.
81. F. Zhu, Z. Shi, C. Qin, L. Tao, X. Liu, F. Xu, L. Zhang, Y. Song, X. Liu, J. Zhang, B. Han, P. Zhang, and Y. Chen, *Nucleic Acids Res* 2012, **40**, D1128-D1136.
82. C. Fu, G. Jin, J. Gao, R. Zhu, E.B. Villagrana, and S.T. Wong, *Bioinformatics* 2013, **29**, 1834-1836.
83. C. Von Mering, M. Huynen, D. Jaeggi, S. Schmidt, P. Bork, and B. Snel, *Nucleic Acids Res* 2003, **31**, 258-61.

Tables: (Two)

Centrality measures	Vulnerability	Closeness	Centroid values	SP betweenness	CF betweenness	Eigen vector
Vulnerability	1.00	0.88	0.85	0.87	0.83	0.85
Closeness	--	1.00	0.94	0.75	0.78	0.98
Centroid values	--	--	1.00	0.81	0.86	0.91
SP betweenness	--	--	--	1.00	0.95	0.71
CF betweenness	--	--	--	--	1.00	0.76
Eigen vector	--	--	--	--	--	1.00

Table 1: The pair wise correlation coefficients between six different centrality measures.

S. no.	Predicted candidate genes	Full name	Gene ID
1	ARRB1	Arrestin, beta 1	408
2	ARRB2	Arrestin, beta 2	409
3	CAV1	Caveolin 1, caveolae protein, 22kDa	857
4	CFTR	Cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7)	1080
5	EP300	E1A binding protein p300	2033
6	ERBB3	v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)	2065
7	HIF1A	hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)	3091
8	INSR	Insulin receptor	3643
9	JAK2	Janus kinase 2	3717
10	JUN	Jun-proto-oncogene	3725
11	LYN	v-yes-1 Yamaguchi sarcoma viral related oncogene homolog	4067
12	PML	promyelocytic leukemia	5371
13	RET	ret proto-oncogene	5979
14	SMAD3	SMAD family member3	4088
15	SRC	v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian)	6714

Table 2: The predicted 15 candidate genes of cervical cancer and their corresponding Gene ids obtained from NCBI database.

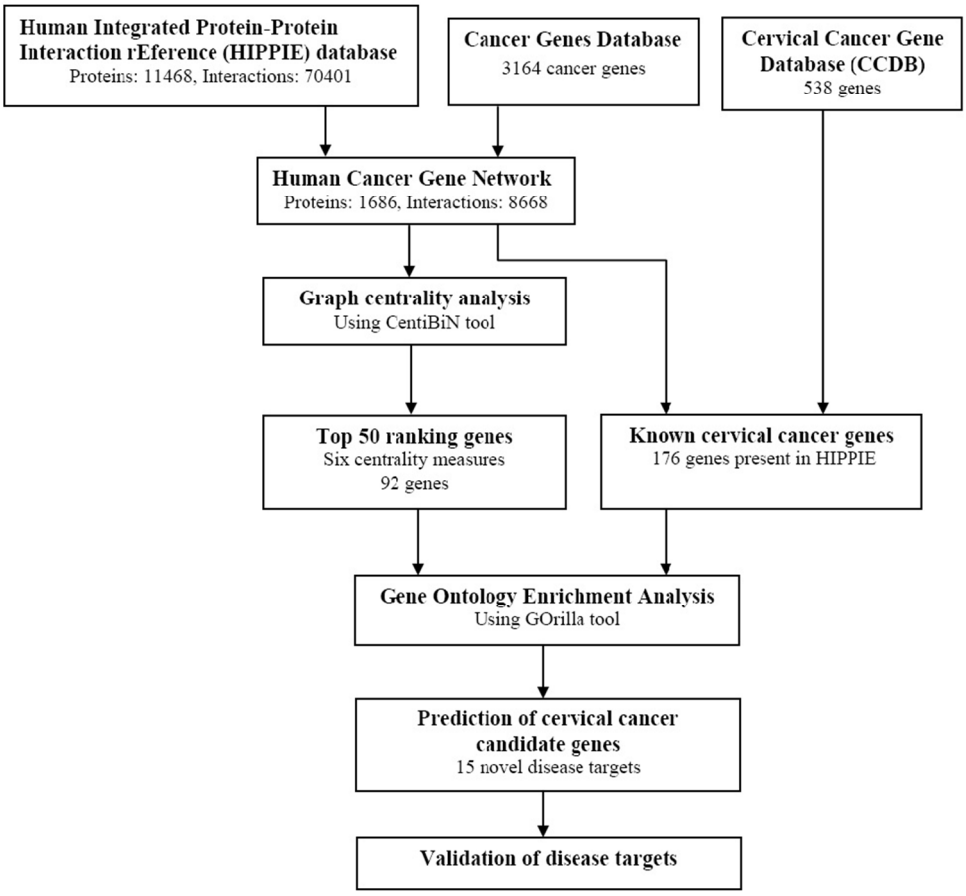
Figures Legends: (Four)

Figure 1: The graphical abstract of the strategy to predict the candidate genes.

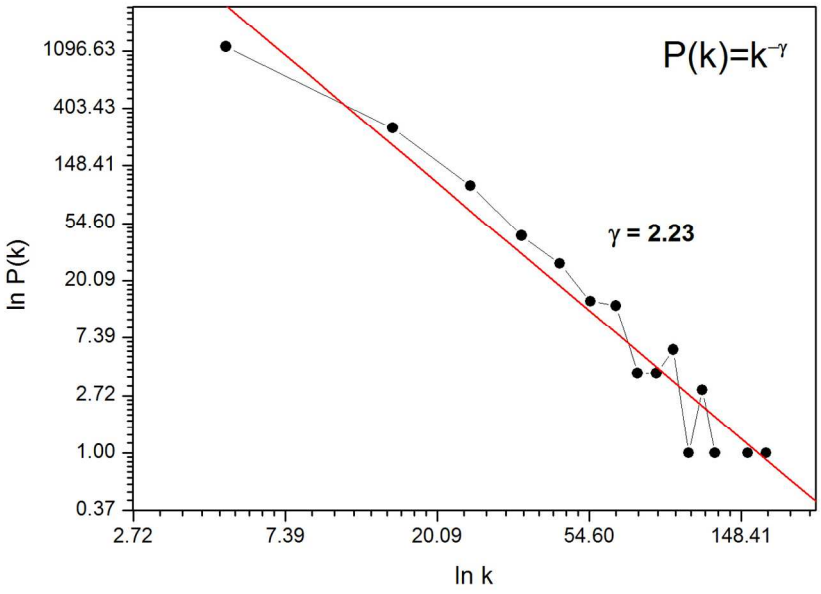
Figure 2: Degree distribution of HCGN observed to follow power law with an exponent □= 2.23 and exhibits the scale free nature.

Figure 3: Venn diagram to predict the candidates for cervical cancer. The pooled list of top 50 ranking genes of each centrality measure, known cervical cancer genes set and the list of genes with significant disease ontologies were logically juxtaposed 15 novel genes were predicted to be candidate genes for cervix related cancer.

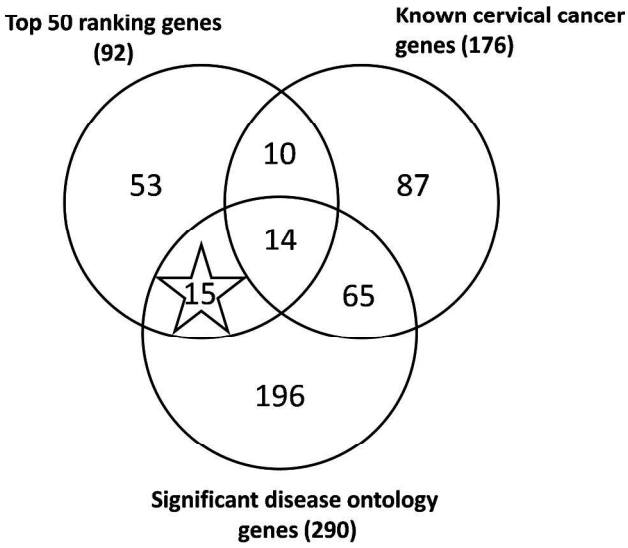
Figure 4: The 15 predicted candidate genes of cervical cancer and their existence in the respective centrality measure, magenta color represents the presence of gene.



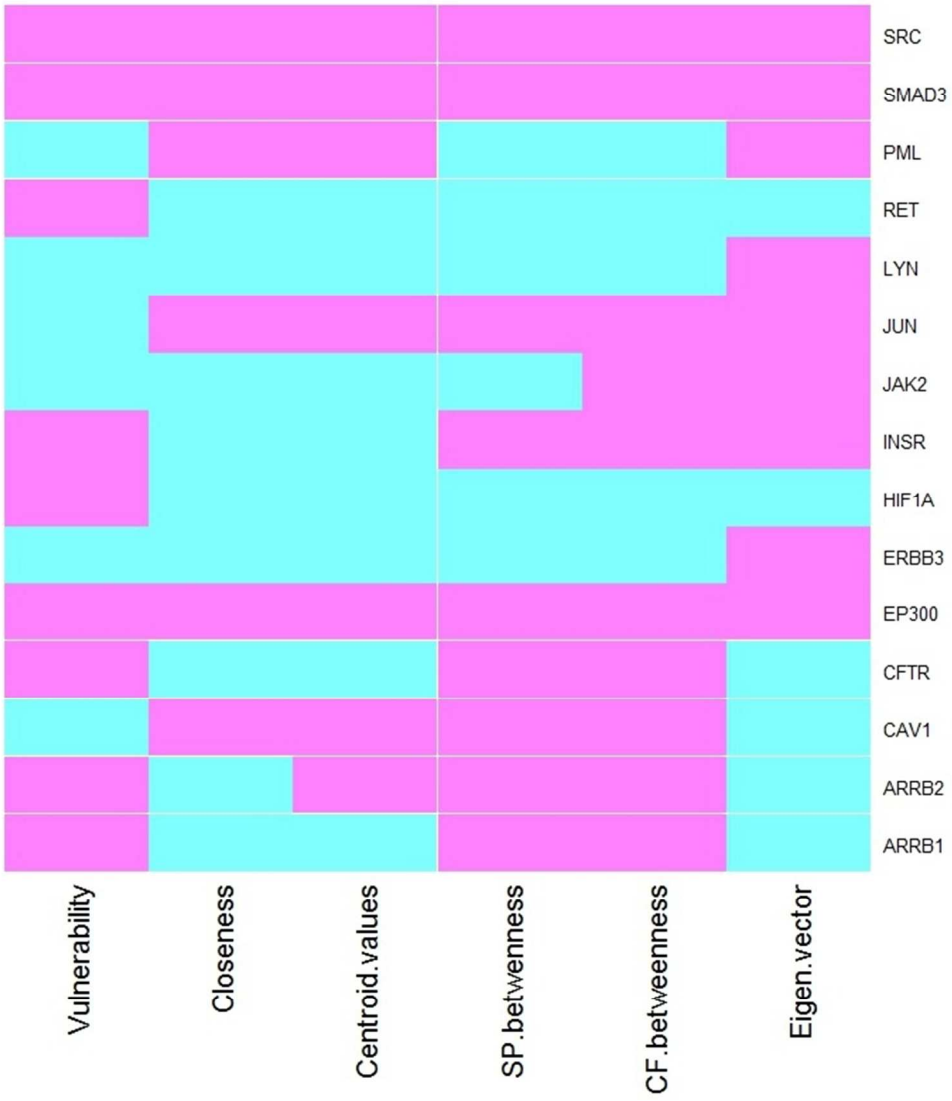
The graphical abstract of the strategy to predict the candidate genes.
269x246mm (96 x 96 DPI)



Degree distribution of HCGN observed to follow power law with an exponent Y= 2.23 and exhibits the scale free nature.
286x201mm (150 x 150 DPI)



Venn diagram to predict the candidates for cervical cancer. The pooled list of top 50 ranking genes of each centrality measure, known cervical cancer genes set and the list of genes with significant disease ontologies were logically juxtaposed 15 novel genes were predicted to be candidate genes for cervix related cancer.
254x190mm (300 x 300 DPI)



The 15 predicted candidate genes of cervical cancer and their existence in the respective centrality measure, magenta color represents the presence of gene.
187x206mm (96 x 96 DPI)