An Approach for the Identification of Targets Specific to Bone Metastasis Using Cancer Genes Interactome and Gene Ontology Analysis

Shikha Vashisht, Ganesh Bagler*

Biotechnology Division, Institute of Himalayan Bioresource Technology, Council of Scientific and Industrial Research, Palampur, India

Abstract

Metastasis is one of the most enigmatic aspects of cancer pathogenesis and is a major cause of cancer-associated mortality. Secondary bone cancer (SBC) is a complex disease caused by metastasis of tumor cells from their primary site and is characterized by intricate interplay of molecular interactions. Identification of targets for multifactorial diseases such as SBC, the most frequent complication of breast and prostate cancers, is a challenge. Towards achieving our aim of identification of targets specific to SBC, we constructed a 'Cancer Genes Network', a representative protein interactome of cancer genes. Using graph theoretical methods, we obtained a set of key genes that are relevant for generic mechanisms of cancers and have a role in biological essentiality. We also compiled a curated dataset of 391 SBC genes from published literature which serves as a basis of ontological correlates of secondary bone cancer. Building on these results, we implement a strategy based on generic cancer genes, SBC genes and gene ontology enrichment method, to obtain a set of targets that are specific to bone metastasis. Through this study, we present an approach for probing one of the major complications in cancers, namely, metastasis. The results on genes that play generic roles in cancer phenotype, obtained by network analysis of 'Cancer Genes Network', have broader implications in understanding the role of molecular regulators in mechanisms of cancers. Specifically, our study provides a set of potential targets that are of ontological and regulatory relevance to secondary bone cancer.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: bagler@ihbt.res.in, ganesh.bagler@gmail.com

Introduction

Cancer is a disease of multiple systems and components that interact at both molecular and cellular levels leading to initiation, progression and spread of the disease [1,2]. The changing interactions of these systems in a dynamic environment underscore the inherent complexity of the disease. Until recently, cancer has been studied with a reductionist approach focusing on a specific mutation or a pathway. Lately there has been a tremendous increase in systems-level study of cancer and the use of integrative approaches to understand mechanisms of cancers [3,4] and their metastases [5,6].

Metastasis is one of the most enigmatic hallmarks of cancers characterized by complex molecular interactions [1,7]. It is responsible for as much as 90% of cancer-associated mortality, yet remains the most poorly understood component of cancer pathogenesis [7,8]. Tumor metastasis is a multistage process during which malignant cells spread from the primary tumor to discontiguous organs [7]. Metastatic dissemination involves a sequence of steps involving invasion, intravasation, extravasation, survival, evasion of host defense and adaptation to the foreign microenvironment [7,8].

Secondary bone cancer (SBC) is a complex disease involving interplay of osteolytic and osetoblastic mechanisms [9] (Figure 1). Bone metastases are the most frequent complication of breast and prostate cancers with a very high propensity of metastasizing to bone causing bone pain, fracture, hypercalcemia and paralysis [10–13]. Breast and prostate carcinomas are often known to take years to develop metastatic colonies (in a limited number of sites) suggesting that in these cancers, cells employ distinct adaptive programs to laboriously cobble together complex shifts in gene-expression programs [14]. Many molecules and associated pathways are reported to be involved in metastasis of cancer cells from breast cancer [13,15–22] and those from prostate cancer [10,18,23–30].

Cancers are characterized by hallmark processes and shared mechanisms involved in expression of disease phenotype. It is a challenge to identify such genes involved in generic cancer mechanisms. Identification of such 'generic cancer genes' may help us focus on 'disease specific cancer genes' of potential therapeutic value. Due to complexity and subtle mechanisms involved in metastasis, it is difficult to identify their control mechanisms. Therefore it is important to have methods for identification of genes and regulatory mechanisms that are key to a complex pathogenic state such as secondary bone cancer. Complex network models of interactomes, along with graph-theoretical analysis and overrepresentation studies, present us a useful strategy for probing molecules that are central to SBC mechanisms and hence potential therapeutic targets.
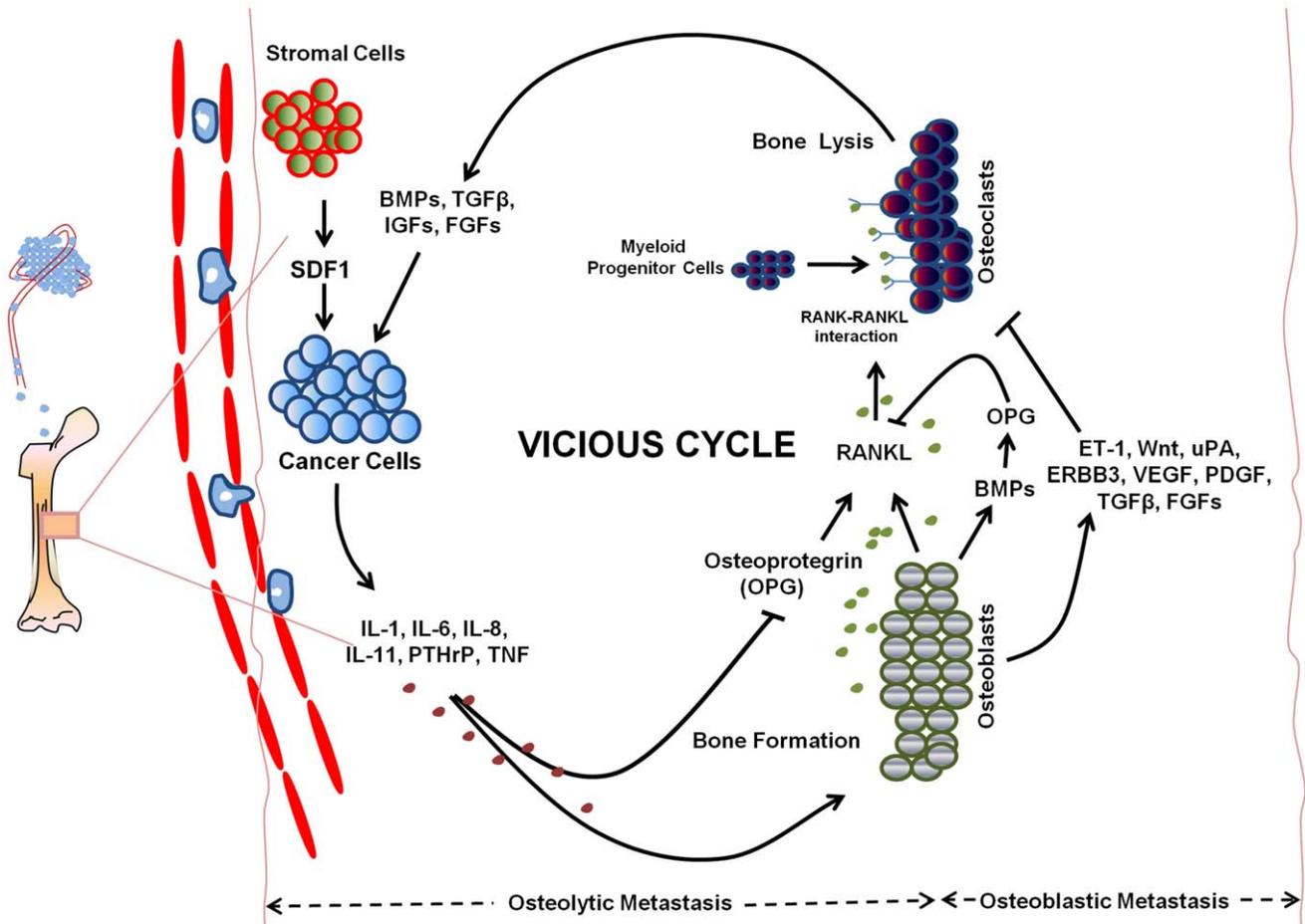
Figure 1. Regulatory mechanisms underlying metastasis to bone reflecting complex interplay of molecules. Bone metastasis results from imbalance of normal bone remodeling process involving osteolytic (leading to bone destruction) and osteoblastic (leading to aberrant bone formation) mechanisms. Breast cancer metastases are usually osteolytic, whereas prostate cancer metastases are usually osteoblastic. **Osteolytic metastasis**: Osteolytic metastasis of tumor cells involves a "vicious cycle" between tumor cells and the skeleton. The vicious cycle is propagated by four contributors: tumor cells, bone-forming osteoblasts, bone resorbing osteoclasts and stored factors within bone matrix. Osteoclast formation and activity are regulated by the osteoblast, adding complexity to the vicious cycle. Tumor cells release certain factors including IL-1, IL-6, IL-8, IL-11, PTHrP and TNF that stimulate osteoclastic bone resorption. These factors enhance the expression of RANKL over OPG by osteoblasts, tipping the balance toward osteoclast activation thus causing bone resorption. This bone lysis stimulates the release of BMPs, TGFβ, IGFs and FGFs for stimulating the growth of metastatic cancer cells to bone. **Osteoblastic metastasis:** Factors released by osetoblastic cells, such as ET-1, Wnt, ERBB3, VEGF play an important role in osteoblastic metastasis by increasing cancer cell proliferation and enhance the effect of other growth factors including PDGF, FGFs, IGF-1. Osteoblast differentiation is also increased by BMPs through the activation of certain transcription factors. Urokinase Plasminogen Activator (uPA), a protease, also acts as mediator for osteoblastic bone metastasis by cleaving osteoclast-mediated bone resorption factors responsible for regulation of osteoclast differentiation; thereby blocking the bone resorption.
doi:10.1371/journal.pone.0049401.g001

Cellular functions reflect the state of the cell as a function of an intricate web of interactions among large number of genes, metabolites, proteins and RNA molecules. A disease phenotype reflects various pathobiological processes that interact in a complex network and is rarely a consequence of abnormality in a single effector gene product [3]. Understanding diseases in the context of organizing principles of the architecture of biological networks allows us to address some fundamental properties of genes that are involved in disease. Study of disease protein interactomes offer a better understanding of disease-specific genes and processes involved and may offer better targets for drug development. Molecular interaction networks are characterized by the presence of a few highly connected nodes, often called hubs, suggesting a special role of these promiscuous interactors. Hubs of protein interactomes are more likely to be essential for the survival [31] and also reported to be important for cellular growth rate

[32]. Proteins with high betweenness [33,34] are reported to have much higher tendency to be essential genes [35,36]. Cancer proteins are reported to be more central in the protein interactome and are, on an average, involved in twice as many interactions as those of non-cancer proteins [37].

The Gene Ontology [38] project provides an ontology of defined terms representing gene product properties. The ontology covers three domains: cellular component, parts of a cell or its extracellular environment; molecular function, elemental activities of a gene product at the molecular level and biological process, sets of molecular events with a defined beginning and end, pertinent to the functioning of integrated living units. GO enrichment methods provide a way to extract biological insight from a set of genes using the power of gene sets [39]. Enrichment analysis involves identification of GO terms that are significantly overrepresented in a given set of genes using statistical models such as

hypergeometric and chi squared distributions [40]. A large repertoire of tools has been developed in recent years for enrichment analysis [41,42]. Methods of network analysis and enrichment studies have been effectively used to identify targets of diseases such as chronic fatigue syndrome [43], major depressive disorder [44], glioblastoma [45], colorectal carcinogenesis [46] and primary immunodeficiency disease [47].

In this study, we aimed to identify secondary bone cancer specific genes. Towards this goal, we used a composite strategy (Figure 2) involving identification of cancer genes specifying generic cancer mechanisms, compilation of genes implicated in metastasis to bone and identification of genes annotated with GO terms specific to secondary bone cancer, to obtain disease specific targets. While network analysis provides a systems perspective of complex molecular mechanisms and helps to identify its central components (functional elements), gene enrichment method enables identification of characteristic ontological features of the gene sets. We first constructed a representative protein inter-actome of all cancer genes and obtained hubs that are involved in generic cancer mechanisms. Further, we compiled a set of experimentally verified genes from the literature that is involved in metastasis of primary breast and prostate cancer into bone, the dominant cause of secondary bone cancer. Using a combination of protein interactome analysis and gene ontology enrichment studies, we obtained a set of genes (targets) specific to SBC mechanisms. Our study provides an approach to identify targets specific to a complex disease phenotype (bone metastasis) by combining systems-level interactome analysis and ontological studies.

Results

CGN as a Representative Interactome of Cancer Mechanisms

We intended to construct an interactome that represents mechanisms involved in processes contributing to cancers. For this purpose we used genes listed in CancerGenes database [48], a compilation of cancer genes that are causally implicated in oncogenesis. We obtained 3164 cancer genes from CancerGenes database, which were used to construct an interactome. These genes were mapped on Human Protein Reference Database [49], a database of curated proteomic information pertaining to human proteins, to construct the Cancer Genes Network (CGN). CGN thus represents an intricate network of cancer proteins. CGN comprises 11602 interactions among 2665 proteins. Figure 3 depicts the CGN, a representative protein interactome of molecular agents and their regulatory interactions, involved in disease phenotype of cancers. Based on earlier reports [3], we hypothesize that proteins that are key to the structural integrity and interaction dynamics of CGN would correspond to proteins involved in regulatory mechanisms generic to cancers.

Topologically Central Genes of CGN Correlate to Generic Cancer Mechanisms

Molecular interaction networks have been reported to have a scale-free nature marked by the presence of few hubs that are critical for the networks [50]. Such hubs of protein interactomes are reported to be more essential for the survival [31] and also important for cellular growth rate [32]. Proteins with high betweenness are reported to have much higher tendency to be essential genes [35,36]. Cancer proteins are reported to be more central in the protein interactome and are, on an average, involved in twice as many interactions as non-cancer proteins [37]. As reported for other biological molecular networks [50,51], we find

that CGN has a scale-free nature indicating presence of exceptionally promiscuously interacting hub nodes and those mediating a large number of interactions (Figure 4) [52].

We computed seven parameters that reflect topological features (degree [53] and neighborhood connectivity [54]), network flow (betweenness [33,34], stress [33,34] and average shortest path length [55]) and local clustering (clustering coefficient [53,55,56] and topological coefficient [57]). The 'key' proteins, that are topologically and dynamically central to CGN, were identified by network analysis. For this purpose 'degree', 'stress' and the normalized counterpart of the latter, 'betweenness', were used for identification of central proteins of CGN. We find that the selected parameters show very high mutual positive correlations (Figure 5). The best-ranked proteins from each of these parameters were compiled, for various thresholds (Top25–Top200), to identify proteins designated as 'hubs' (Set-A in Figure 2 and Table S1).

We find that the 'hubs' (Figure 3; Top75 threshold), thus identified, include genes involved in regulation of processes known to be generically present in most cancers [1]. Proteins known to be involved in growth signals, thus leading to self-sufficiency in cancer cells, such as TGFBR1, EGFR, IGFR1R, GRB2, are in the 'hubs' of CGN [2]. The hubs also include RB1, BCL2, AKT1, CDK2, CASP8 which are involved in mechanisms of apoptosis, to which cells are known to acquire resistance in all types of cancers [1,58]. The TP53 tumor suppressor protein, known to be involved in most commonly occurring loss of proapoptotic regulation and affecting the apoptotic effector cascade [59,60], is also present in the CGN hubs. One of the hub proteins, SMAD4, is known to be involved in differentiation, apoptosis and cell cycle [61]. MYC, known to upregulate cyclins and downregulate CDKN1A (P21), is also one of the hubs of CGN.

Further, we checked how well the identified hubs correlate with KEGG-PIC [62], a collection of genes from generic pathways involved in cancers (Table S2). Figure 6 shows the overlap of hub genes (Top25 to Top200) with KEGG-PIC genes. We find that indeed the Top25 hubs, comprising the most promiscuously interacting proteins and mediating a large number of interactions in cancer genes interactome, have a 74% overlap with KEGG-PIC. The precision of identification of generic cancer proteins drops as the definition of 'hubs' is relaxed further (57% for Top50 and 52% for Top75). As a negative control, for the ability of network-metrics-based ranking to identify generic proteins, we use genes ranked worst according to the selected parameters. We find that these genes have a very poor overlap with KEGG-PIC genes, even when compared to that expected from corresponding random samplings (Figure 6). We treat Top25, Top50 and Top75 CGN hubs as representatives of generic cancer genes and use them, further, to identify SBC-specific targets.

GO Enrichment of CGN Hub Genes

We performed GO enrichment analysis of CGN 'hub' genes identified. We expect the 'hubs' to represent generic cancer processes which are shared by most, and perhaps, all types of human cancers. The molecular machinery regulating proliferation, differentiation and death of all mammalian cells is highly similar [1]. The genetic transformation of normal body cells results in defects of regulatory circuits that govern normal cell proliferation and homeostasis, and collectively dictate malignant growth.

Cancer cells show self-sufficiency of growth signals involving alteration of transcellular or intracellular mechanisms [2]. The growth signaling pathways are suspected to suffer deregulation in all human tumors [2]. This was reflected in the following significantly overrepresented GO terms of CGN hubs: cellular response to growth hormone stimulus (GO:0071378), regulation of
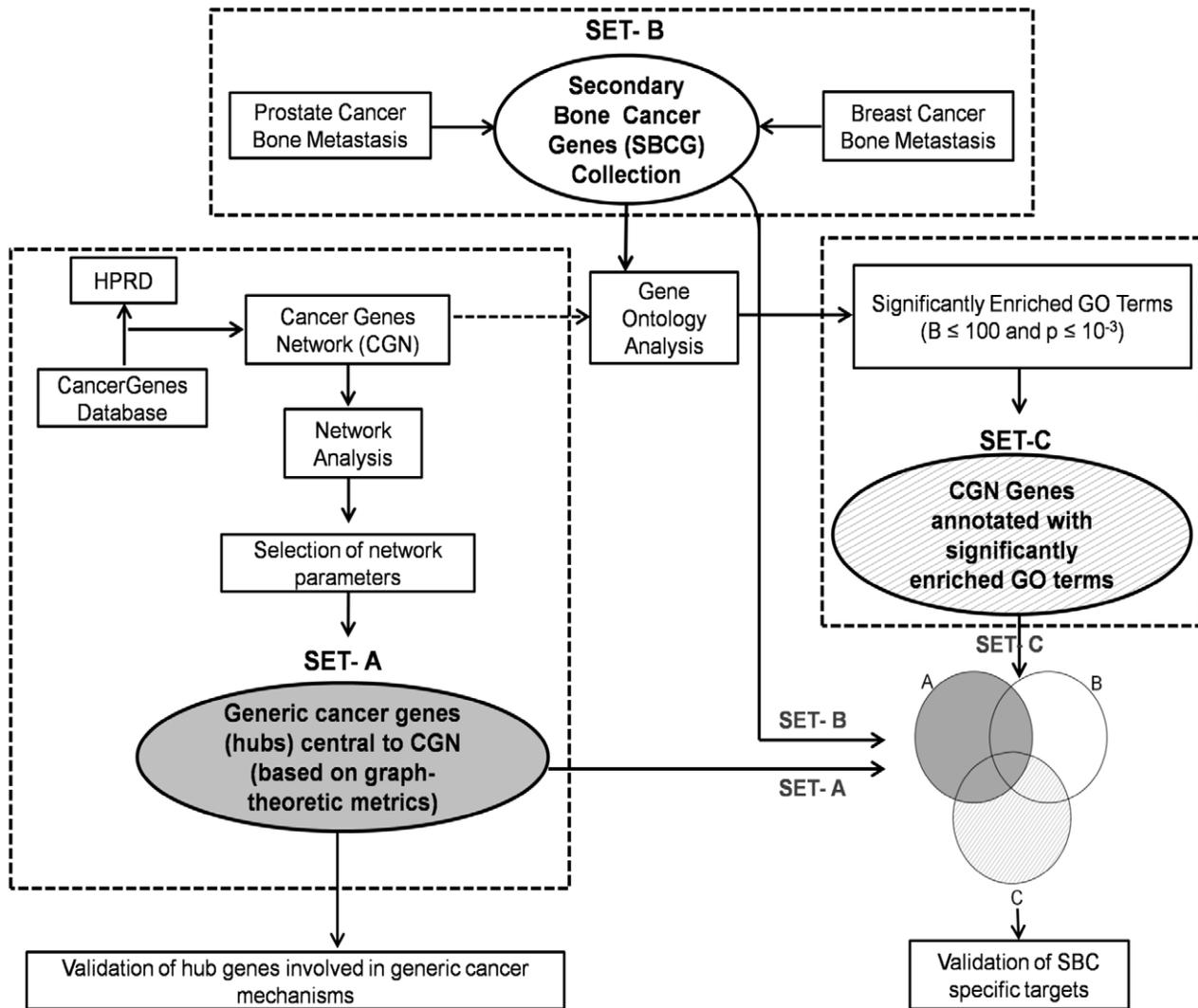
Figure 2. Strategy implemented for the identification of targets specific to secondary bone cancer. The strategy that was implemented in this work comprised three tasks, leading to three corresponding gene-sets that were used for obtaining SBC-specific targets of potential therapeutic value. (a) A compilation of cancer genes from CancerGenes database was used to construct a representative cancer genes interactome (Cancer Genes Network; CGN) by mapping them on to a reference human protein interactome (Human Protein Reference Database; HPRD). Using methods of network analysis, proteins that are central to CGN and interaction dynamics were obtained. These set of genes (SET-A; shaded area) was found to be correlating well with genes implicated in generic cancer mechanisms (Figure 6) as well as those annotated as essential using mouse phenotype data (Figure 8). The CGN, comprising of 11602 interactions among 2665 proteins, also serves as a reference set (universe) for gene enrichment studies; (b) A set of genes (Secondary Bone Cancer Genes; SBCGs) that are implicated in metastasis to bone from primary breast and prostate cancer, the most prevalent causes of bone metastasis, was compiled from literature. This set (SET-B) serves as a basis of genes and ontological correlates of secondary bone cancer that characterize the disease phenotype; (c) Significantly enriched GO terms that characterize SBCGs were obtained by overrepresentation analysis against the 'cancer genes' universe. SET-C, a subset of CGN, was obtained by segregating cancer genes that were annotated with these SBC-specific ontological terms. Part of SET-C (hatched area; Set-c and Set-bc in Figure 10A) serves as a 'source set of target cancer genes' that, both, carry ontological essence of SBCGs and are not involved in generic cancer mechanisms. SBC-specific targets (Figure 3 and Figure 10B), that are annotated with key GO terms (Figure 9) reflecting role in, both, bone processes and metastasis mechanisms, were further obtained from the source set.
doi:10.1371/journal.pone.0049401.g002

epidermal growth factor receptor signaling pathway (GO:0042058), transforming growth factor beta receptor signaling pathway (GO:0007179), signal complex assembly (GO:0007172).

In contrast to normal cells, which respond to antigrowth signals, cancerous cells are insensitive to antigrowth signals due to disruption of pathways related to cell cycle clock. We find that our list of significantly enriched GO terms for CGN hubs had many terms associated with cell cycle processes: mitotic cell cycle G1/S transition checkpoint (GO:0031575), regulation of G1/S

transition of mitotic cell cycle (GO:2000045), DNA damage response, signal transduction by p53 class mediator (GO:0030330), regulation of stress-activated MAPK cascade (GO:0007173).

In human carcinogenesis, all types of cancers are known to acquire resistance to apoptosis [58]. The overrepresented GO terms of CGN hubs underline this observation: release of cytochrome c from mitochondria (GO:0001836), activation of pro-apoptotic gene products (GO:0008633), cell-type specific
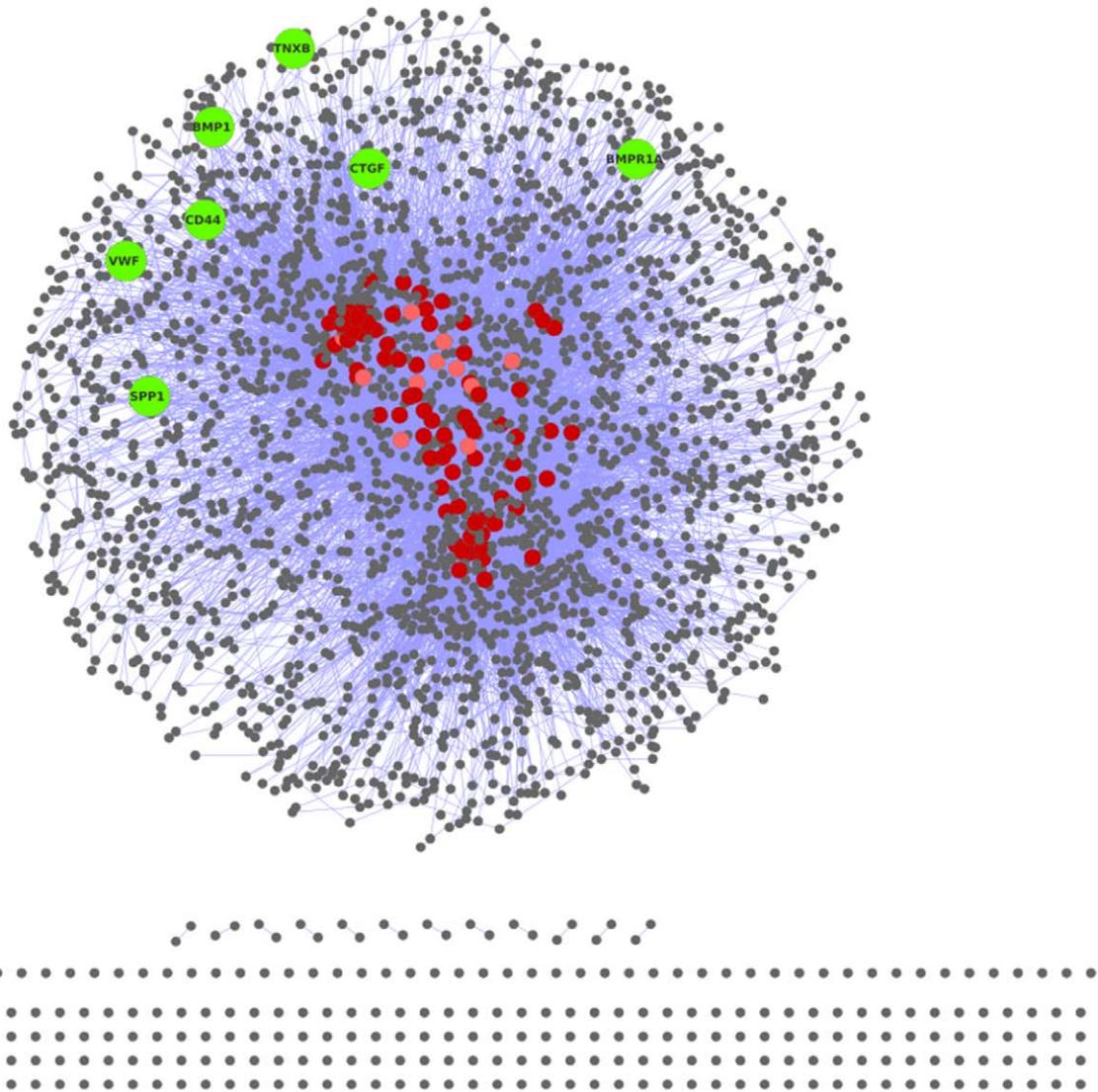
Figure 3. CGN, a representative protein interactome of cancer genes, Top75 hub genes and SBC-specific targets. Cancer Genes Network (CGN) is a protein interactome of cancer genes embodying molecular mechanisms of cancers. Each node represents a cancer protein and an edge between two nodes represents a protein-protein interaction. The giant cluster comprises 89% of all cancer genes. The hubs of CGN, cancer genes central to the structural stability and information dynamics, are depicted in shades of red. The hubs encode genes involved in generic cancer mechanisms and those classified as essential using phenotypic data from Mouse Genome Informatics. 88% of these hubs are essential ('dark red') and the rest non-essential ('light red'). SBC-specific targets, with ontological role in bone and metastasis processes, are highlighted in 'green'.
doi:10.1371/journal.pone.0049401.g003

apoptotic process (GO:0097285), cellular component disassembly involved in apoptotic process (GO:0006921), positive regulation of anti-apoptosis (GO:0045768).

In addition, we find following significantly enriched 'molecular function' terms: transforming growth factor beta receptor, pathway-specific cytoplasmic mediator activity (GO:0030618), receptor signaling protein tyrosine kinase activity (GO:0004716), ephrin receptor binding (GO:0046875), MAP kinase activity (GO:0004707), beta-catenin binding (GO:0008013), cytokine receptor binding (GO:0005126), androgen receptor binding (GO:0050681).

All the GO-terms mentioned above broadly support generic cancer processes. The p-values for these 'significantly enriched GO terms of CGN Top75 hubs' are in the range of 10e-4 and 10e-13. Figure 7 depicts these overrepresented GO terms and statistics of their prevalence.

Topologically Central Genes of CGN Correlate with Biological Essentiality

We further checked the biological relevance of the central genes of CGN. We divided CGN into two sets, essential and non-essential, using Mammalian Phenotype Ontologies [63] to classify genes as essential when they caused embryonic, perinatal, postnatal or neonatal lethality in mouse models [64] using phenotypic data from Mouse Genome Informatics (MGI) [65]. The percentage of essential genes were computed in the 'hubs' and their corresponding non-hubs for varying hub definitions. Figure 8 shows the statistics for percentage of hub and non-hub genes annotated as essential. Among the seven parameters computed, the hubs identified using degree, betweenness and stress have a significant percentage of essential genes (between 82% to 92%) (Figure 8A). Incidentally, none of the remaining four network
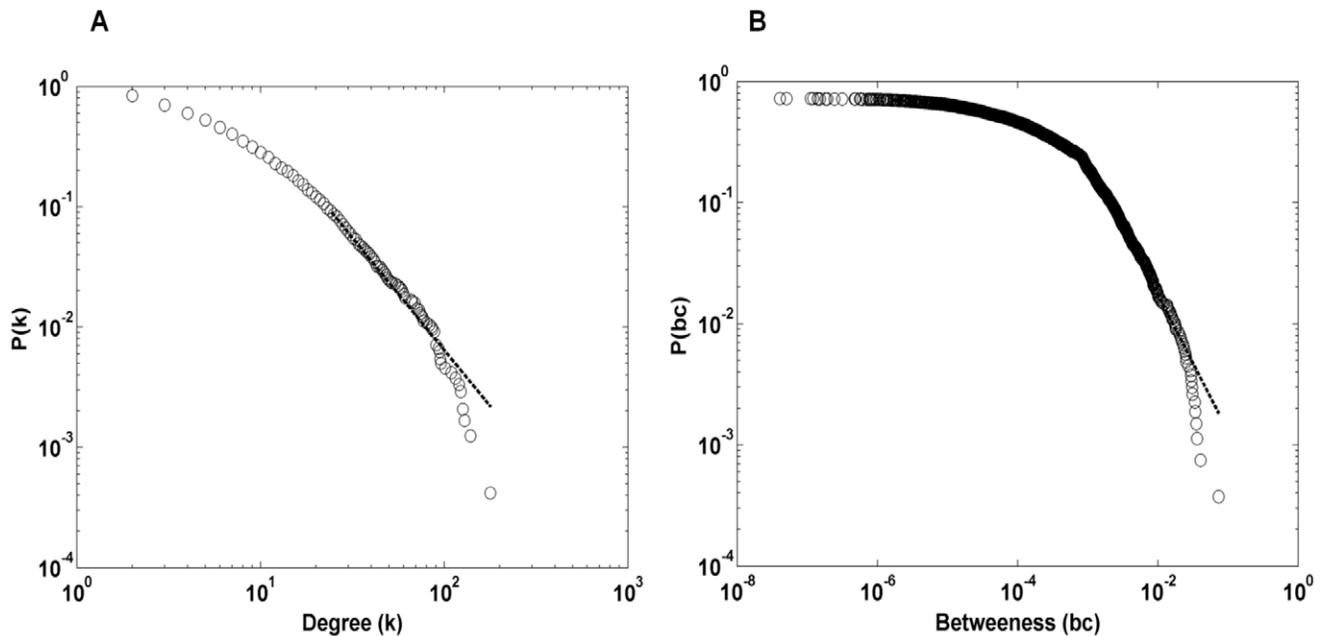
Figure 4. Scale-free nature of degree and betweenness distributions of CGN. The distributions of (A) degree and (B) stress (as well as its normalized counterpart 'betweenness') show a scale-free nature. Dotted lines show the power law fit with an exponent of -2.85 and -2.13, respectively. This indicates presence of promiscuously interacting proteins with high degrees and central mediators that play role in information dynamics across the network, respectively.
doi:10.1371/journal.pone.0049401.g004

parameters show good correlation with biological essentiality. The non-hubs, as expected from the random samplings statistics, have around 50% of essential genes associated, regardless of the metric used for ranking (Figure 8B). For hubs comprising of Top75 genes, the chosen parameters have a very high proportion of essential genes (between 88% and 89%), compared to the rest of the parameters. Figure 3 depicts the central nature of Top75 hub genes of CGN and highlights the essential and non-essential genes among them. This emphasizes the relevance of the network metrics chosen in encoding biological relevance. This is consistent with earlier reports about hubs being more essential [64] and network centrality of cancer genes [37].

SBCGs and their Characteristic GO Terms

Knowing that metastasis into bone is primarily caused by spread from primary prostate and breast cancers [10–13], we aimed at compiling genes linked with these processes. We compiled a list of 391 Secondary Bone Cancer (SBC) susceptible genes from literature reporting experimental studies (Set-B in Figure 2 and Table S3). For identifying the relevance of these genes in the context of metastasis into bone, we performed an overrepresentation analysis using CGN as universe. Gene ontology enrichment studies of SBCGs are expected to reflect GO categories of biological processes and molecular functions that are relevant for metastasis into bone, consistent with the known aspects of regulatory mechanisms of bone metastasis (Figure 1).

We identified 93 significantly enriched GO terms of SBCGs for biological processes, molecular functions and cellular components. We identify CGN genes annotated with either of these 93 GO terms to construct 'Enriched Genes' set (Set-C in Figure 2). We find that out of the 93 significantly enriched GO terms, 31 are associated with metastasis or bone processes. Out of these 31 GO terms, 21 are most relevant for metastasis mechanisms and 10 for bone-related processes (Figure 9).

The following GO terms, that were significantly enriched, are related to bone processes: osteoblast differentiation (GO:0001649), regulation of bone remodeling (GO:0046850), endochondral ossification (GO:0001958), replacement ossification (GO:0036075), bone mineralization (GO:0030282), ossification (GO:0001503), cartilage development (GO:0051216), response to vitamin D (GO:0033280), collagen metabolic process (GO:0032963), collagen fibril organization (GO:0030199).

The following biological processes GO terms, that were significantly enriched, are related to mechanisms of metastasis: leukocyte migration (GO:0050900), cell-cell adhesion (GO:0016337), angiogenesis (GO:0001525), positive regulation of cell adhesion (GO:0045785), negative regulation of cell adhesion (GO:0007162), positive regulation of leukocyte migration (GO:0002687), regulation of leukocyte chemotaxis (GO:0002688), positive regulation of leukocyte chemotaxis (GO:0002690), positive regulation of catenin import into nucleus (GO:0035413).

Following molecular functions GO terms that were significantly enriched, are related to mechanisms of metastasis: laminin binding (GO:0043236), fibronectin binding (GO:0001968), fibroblast growth factor receptor binding (GO:0005104), platelet-derived growth factor binding (GO:0048407), extracellular matrix binding (GO:0050840), collagen binding (GO:0005518), integrin binding (GO:0005178), glycosaminoglycan binding (GO:0005539), cytokine activity (GO:0005125), carbohydrate binding (GO:0030246).

Predicted SBC-specific Targets

For predicting SBC-specific targets of potential therapeutic value, we logically juxtaposed the three sets of genes obtained: CGN hub genes (Top75), SBCGs and the cancer genes annotated with GO terms overrepresented for SBCGs (Figure 2 and Figure 10). On the basis of specificity for secondary bone cancer, as depicted in Figure 10A, the Venn diagram corresponding to these gene sets was divided into seven distinct regions: Set-a, Set-b,
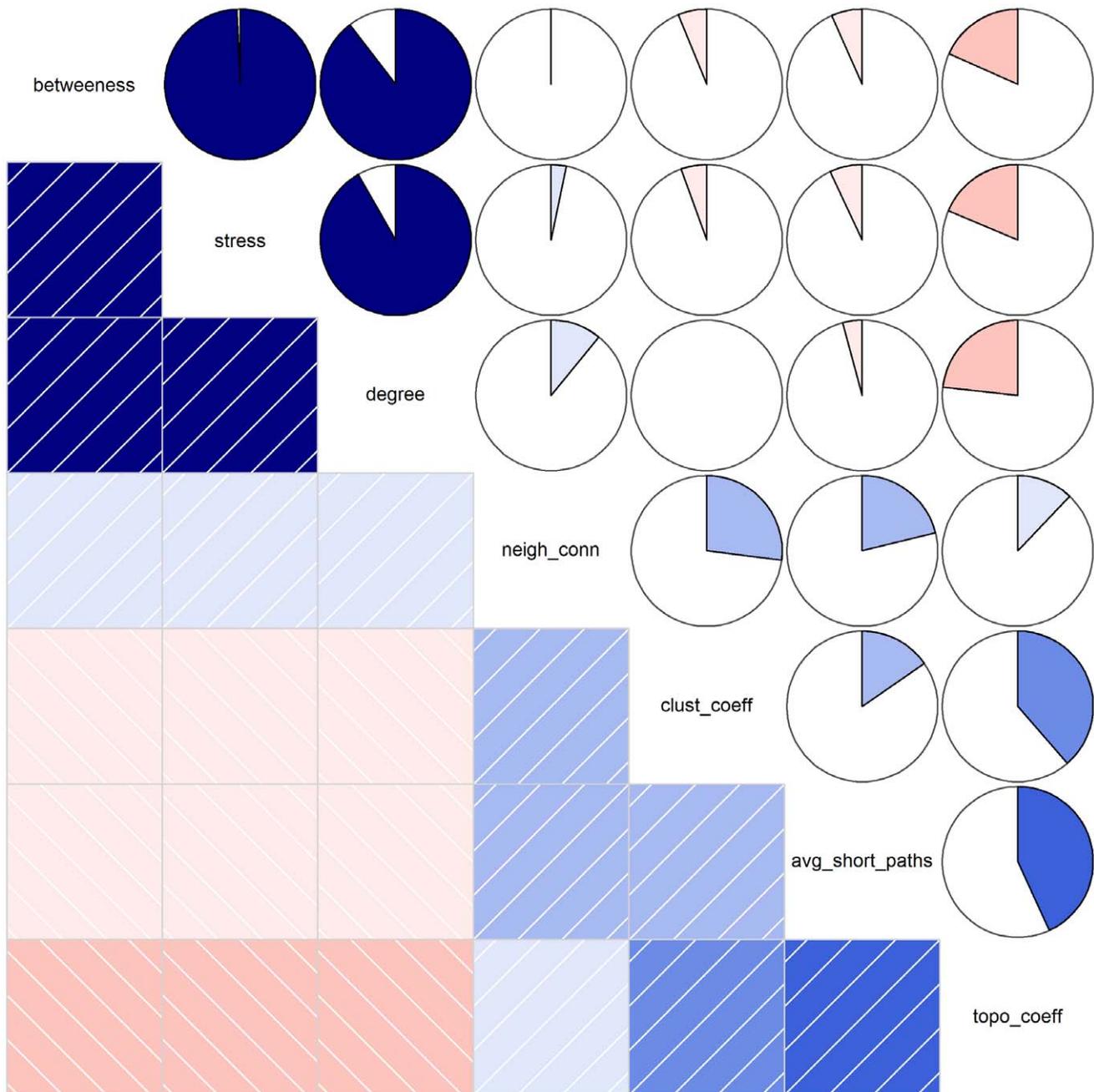
Figure 5. Heatmap of topological metrics computed for Cancer Genes Network. The heatmap of pair-wise correlations among seven parameters that enumerate topological, dynamical and local clustering features of the network: betweenness, stress, degree, neighborhood connectivity (neigh_conn), clustering coefficient (clust_coeff), average shortest path length (avg_short_paths) and topological coefficient (topo_coeff). The heatmap highlights three parameters with very high mutual positive correlations (r = 0.8989, 0.9930 and 0.9210): degree, betweenness and stress. The upper triangle of the heatmap depicts pair-wise correlations as pie charts. The lower triangle depicts positive and negative correlations in shades of blue and red, respectively; the darker the color the stronger the correlation. Positive and negative correlations are also depicted with right- and left-handed diagonal lines.
doi:10.1371/journal.pone.0049401.g005

Set-c, Set-ab, Set-bc, Set-ac and Set-abc. Overall, the sets belonging to central genes of CGN (Set-a, Set-ab, Set-ac and Set-abc) contain genes that are involved in generic cancer mechanisms (Figure 6) and are also essential (Figure 8). Hence, while obtaining SBC-specific targets, we don't consider these gene sets. Out of the rest of three sets, Set-b comprises genes that are already reported to be associated with SBC mechanisms, hence

may not reveal novel genes. We focus on Set-c and Set-bc, corresponding to non-generic cancer genes annotated with SBC-specific GO terms and those reported to be having a role in SBC mechanisms, respectively, for the search of novel SBC-specific targets.

Set-bc and Set-c contain 53 and 134 genes, respectively. These lists were refined to select only those genes that are annotated with
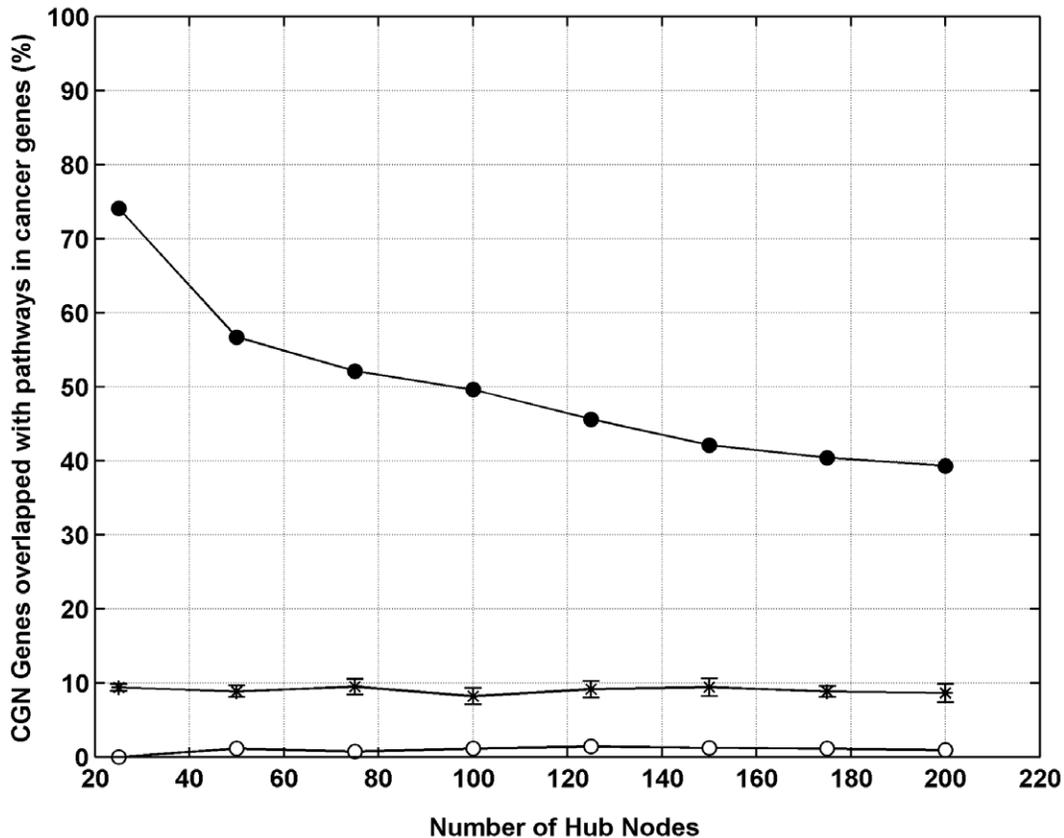
Figure 6. Hubs of CGN correlate with generic cancer genes (KEGG-PIC). The hubs of CGN, identified using chosen network centrality parameters, correlate with the set of generic cancer genes (KEGG-PIC). For hub definitions varying between Top25–Top200, the CGN hubs (top ranked cancer genes) show good overlap with KEGG-PIC genes (filled circles). The correspondence of network centrality and generic role in cancers, expectedly, drops as the strictness of criterion used for identifying hubs is loosened. For the corresponding random samples, the overlap with KEGG-PIC is as expected (stars; error bars indicate standard errors from 1000 samples). The cancer genes with worst centrality rankings (open circles), show almost no overlap with generic cancer genes; worse than that expected from random samplings. Top75, Top50 and Top25 hubs, with 52%, 57% and 74% overlap with generic cancer genes, respectively, were further used for identification of SBC-specific cancer genes.
doi:10.1371/journal.pone.0049401.g006

GO terms most relevant for SBC (Figure 9 and Table S4), to obtain 28 and 60 genes from Set-bc and Set-c, respectively. We find that, in these sets, we were still left with KEGG-PIC genes known to be involved in generic cancer pathways. The list of targets was refined further to obtain a total of 72 targets that are non-generic to cancers and relevant for secondary bone cancer.

Out of these 72 genes, 14 genes are annotated with GO terms relevant to bone processes, 51 with metastasis and 7 with both (Figure 3). We predict these seven proteins, with ontological relevance to bone processes as well as metastasis, to be most potent targets and key regulators of metastasis into bone. The non-generic SBC-specific targets, with relevance to metastasis and bone mechanisms, identified are: SPP1 (Secreted Phospoprotein 1), CD44 (Cluster of Differentiation 44), CTGF (Connective Tissue Growth Factor), TNXB (Tenascin X), BMP1 (Bone Morphogenetic Protein 1), BMPR1A (Bone Morphogenetic Protein Receptor, Type IA) and VWF (Von Willebrand Factor).

We find that the SBC-specific targets identified are indeed relevant for the metastasis into bone and involved in regulation of SBC mechanisms (Table S5). SPP1 is reported to interact with CD44 receptor and is thought to exert pro-metastatic effects leading to tumor progression by regulating the cell signaling events [66,67]. CD44 is also reported to enhance integrin-mediated adhesion and transendothelial migration of breast cancer cells

[68]. CTGF expression has been shown to be associated with tumor development and progession. The overexpression of CTGF in breast cancer cells may promote their metastasis to bone [69]. BMPR1A, belonging to the family of transmembrane serine/threonine kinases, is reported to be necessary for extracellular matrix deposition by osteoblasts, while not essential for osteoblast formation or proliferation [70]. BMP1, which does not belong to TGFβ superfamily unlike other BMPs, is known to induce bone and cartilage development. TNXB is an extracellular matrix glycoprotein expressed in connective tissues including skin, joints and muscles and in known to have role in cell-matrix and cell-cell adhesion. It has been reported that TNX deficiency leads to the invasion and metastasis of tumor cells by facilitating increase in the activity of MMPs which results in the degradation of laminin. The over-expression of TNX could be of potential therapeutic benefit in reducing tumor progression [71]. VWF is a large multimeric glycoprotein present in blood plasma and is produced constitutively in endothelium, megakaryocytes and subendothelial connective tissue. It has been reported that in osteosarcoma tumors the expression of VWF gets deregulated, potentially leading to metastasis [72].
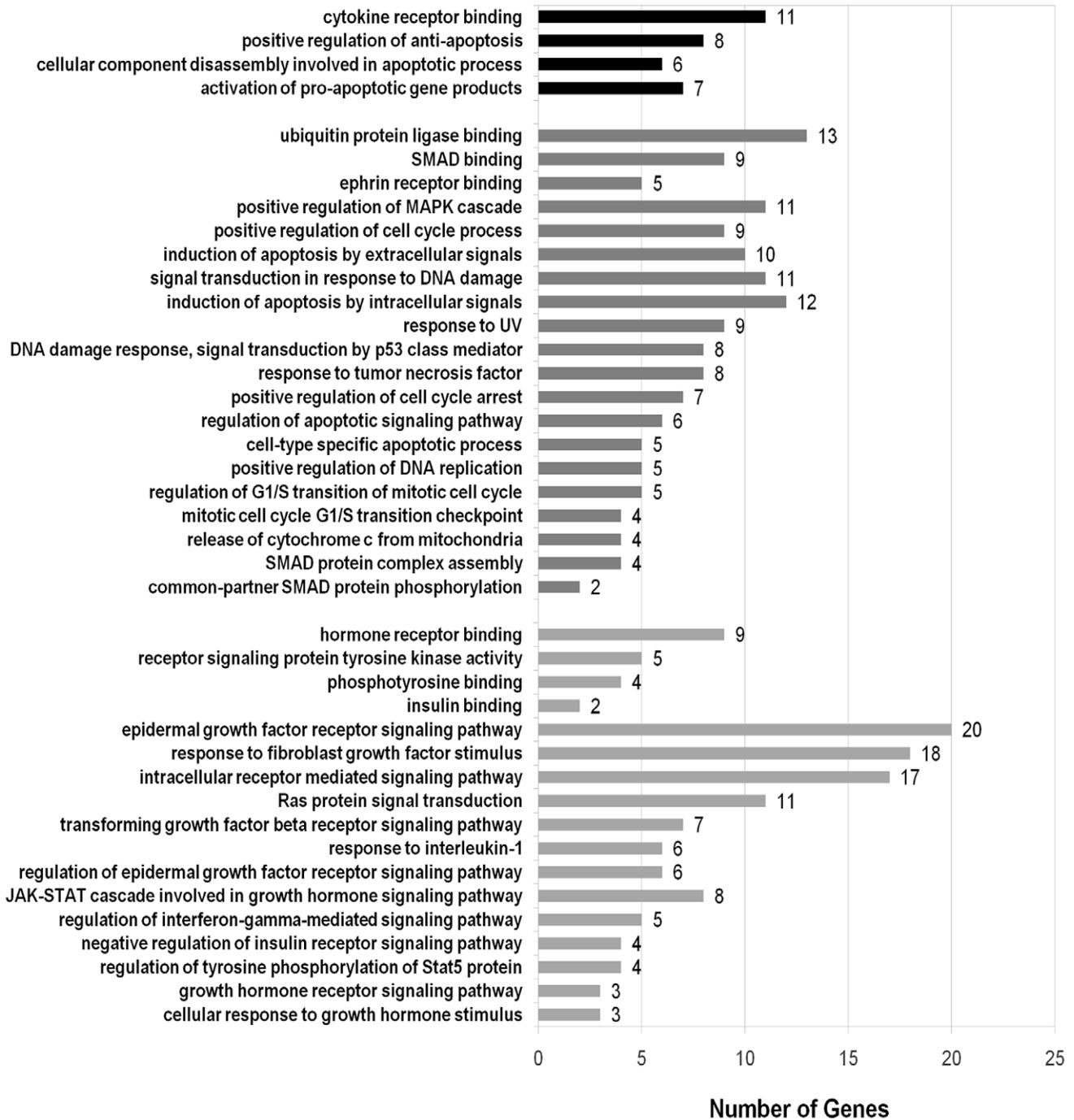
Figure 7. Significantly enriched GO terms for hubs of CGN reflecting their role in generic cancer mechanisms. The overrepresentation studies of Top75 CGN hubs reflect their role in generic cancer processes and functions. Among the significantly enriched GO terms, 17 represent self-proliferation circuits (light-gray), 20 represent cytostasis and differentiation circuits (dark gray) and 4 represent viability circuits (black).
doi:10.1371/journal.pone.0049401.g007

Discussion

Towards our goal of identifying key targets specific for mechanisms of metastases of primary breast and prostate cancers to bone, we used network analysis and gene ontology enrichment studies. The motivation behind using a combination of network analysis and gene enrichment methods is that, while the prior provides a systems perspective of complex molecular mechanisms and helps to identify its central components (functional elements),

the latter enables identification of characteristic ontological features of the gene sets. Secondary Bone Cancer (SBC) is a complex disease triggered from the primary form of cancers, most commonly through breast cancer and prostate cancer. Many pathways and molecular regulators involved in the metastasis of primary prostate cancer and breast cancer are well known. Starting from SBC-related proteins and interactome of proteins involved in cancer mechanisms, using network analysis and
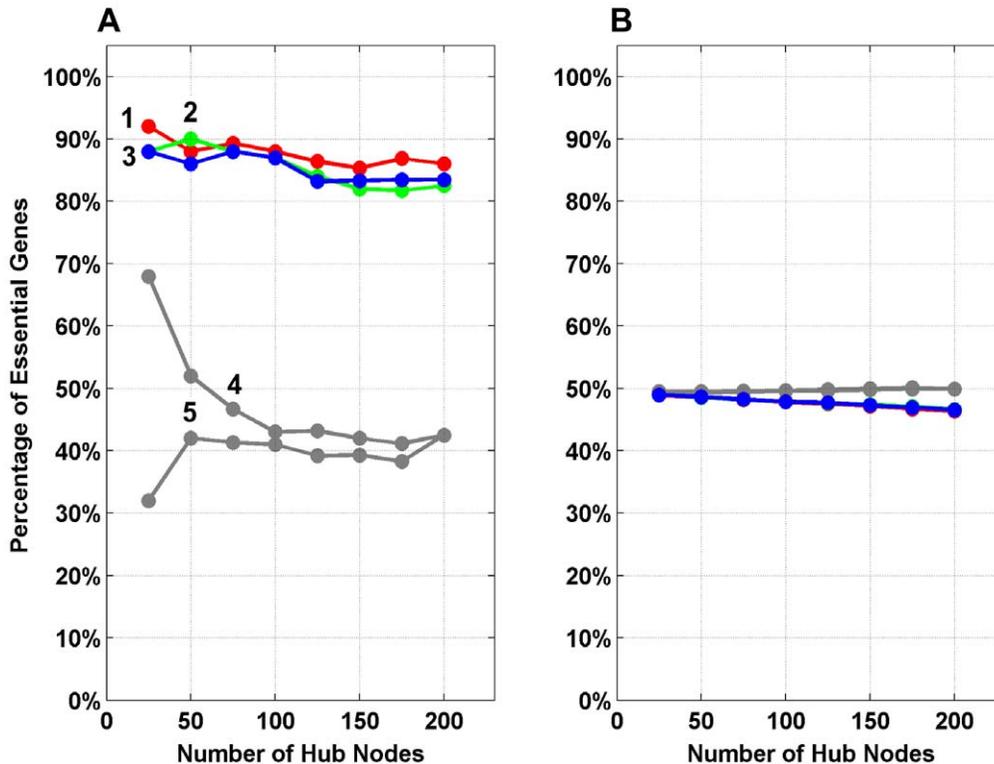
Figure 8. The percentage of essential genes in the hubs and the corresponding non-hubs of CGN. A. For hub definitions varying between Top25 to Top200, the hubs identified using degree (1, 'red'), betweenness (2, 'green'), and stress (3, 'blue') have significantly high percentage (82%–92%) of essential genes, classified using phenotypic data for Mouse Genome Informatics. Among the rest of the four parameters neighborhood connectivity (4) and topological coefficient (5) show neither significant nor consistent correlation with essential genes. Due to the nature of 'clustering coefficient' and 'average shortest path length' parameters, the data could not be binned at the same intervals. For clustering coefficient, percentage of essential genes among the nodes having up to Top200 rankings is in the range of 27% and 62%. For average shortest path length, the percentage of essential genes for nodes having ranking up to Top200 is 26%, worse than expected from random sampling. B. In the corresponding non-hubs, the percentage of essential genes is as expected from random sampling, regardless of the centrality measure or cut-off used for hub definition.
doi:10.1371/journal.pone.0049401.g008

overrepresentation studies, we identify targets that are specific to SBC.

The final list of seven targets was identified using generic cancer genes obtained with Top75 hubs. We repeated this task with Top25 and Top50 lists, which have better overlap with KEGG-PIC genes (74% and 57%, respectively), as compared to that of Top75 hubs (52%) (Figure 7). Using the procedure described, we identified SBC-specific targets starting from Top25 and Top50 CGN hubs. Figure 11 illustrates the results of these experiments obtained with better overlap with generic cancer genes (KEGG-PIC). Using these thresholds we obtained 31 and 60 hubs, respectively, which were filtered from the potential target set. This resulted in increase in number of potential targets in Set-c to 141 and 139, respectively, compared to those obtained with Top75 hubs (134). There was no change in the number of genes (53) obtained in Set-bc. Even with these criteria that correspond better with generic genes, the number of targets (that are annotated with GO terms relevant to metastasis and that of bone processes) does not increase, indicating the soundness of criteria used for identification of SBC-specific targets.

Network analysis has been shown to be a very useful and potent tool in understanding the disease phenotype and probing for therapeutic targets [3]. The identification of network parameters relevant to the question(s) being asked is an important task. We chose degree, betweenness and stress that are known to be

important in the topology of the disease interactome and dynamic interplay of proteins involved [3]. These parameters were also found to have very good correlation with 'essential genes' (Figure 8). The significant terms emerging from overrepresentation analysis of CGN too support our choice of parameters (Figure 7).

Gene Ontology (GO) defines a set of functional terms related by parenthood relationships forming a directed acyclic graph. It produces sets of explicitly defined, structured vocabularies that describe biological processes, molecular functions and cellular components of gene products [38]. GO classification is expected to become an increasingly powerful tool for data analysis and functional predictions as the ontologies and annotations continue to evolve [40]. It is characterized by high quality manual curation, consistent annotation standards across species, and has the advantage of lesser bias as compared to domain specific classification schemes due to its comprehensive nature [40]. GO enrichment methods, which provide means of identification of significantly overrepresented GO terms, could be effectively used to get biological insights from a given set of genes [40]. One of the critical points in GO enrichment analysis is the selection of background set for getting the correct results. In this study, we use a pool of cancer genes (CGN) as universe, which serves the purpose of a meaningful background set of cancer mechanisms.

Cancer is a complex genetic disease characterized by intricate regulations among a diverse set of cancer genes. Hence, it is useful
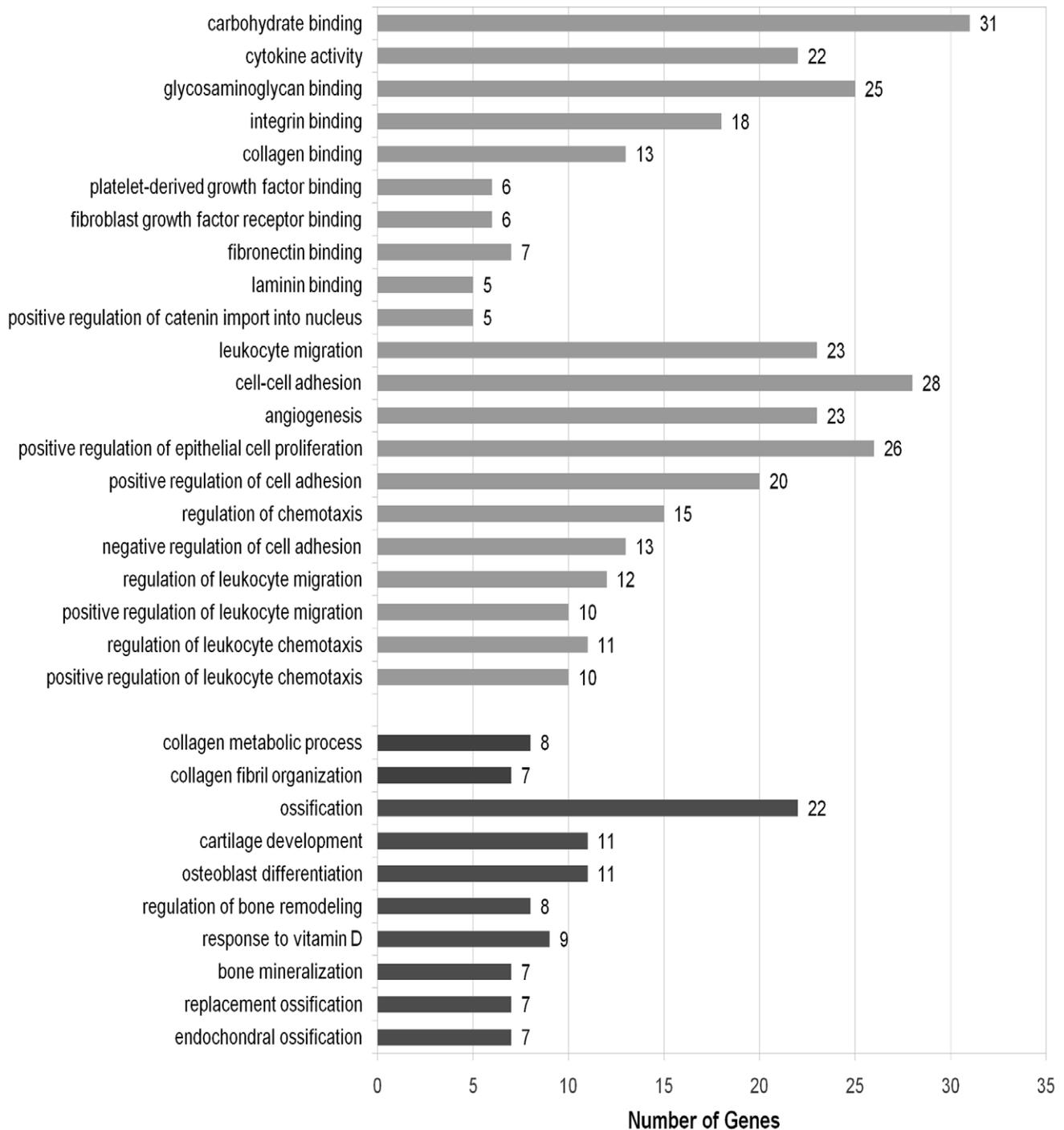
Figure 9. Overrepresented GO terms of processes relevant and necessary for execution of bone metastasis. From the GO enrichment studies of SBCGs, curated from literature, 21 GO terms relevant for metastasis ('light gray') and 10 GO terms relevant for bone ('dark gray') processes were identified.
doi:10.1371/journal.pone.0049401.g009

to have a detailed map of interactions that could help in probing hallmark regulatory modules and perhaps cancer-specific motifs. Here, we used CancerGenes Database (CGDb) [48] as a premise of cancer genes to construct a representative cancer interactome. This could potentially be improved by compiling a more exhaustive list of cancer genes. Though, we believe that the ultimate set of generic cancer genes, central to the topology of the cancer genes network, would not alter much, it may enrich the data and enable a more meaningful analysis of subtle regulatory features of cancer phenotype. It would be interesting to see whether and how the modules of CGN reflect the hallmarks of cancers. Also, while a few of the network centrality metrics for cancer genes may enumerate biological essentiality, it would be interesting to explore a larger parameter space to search for
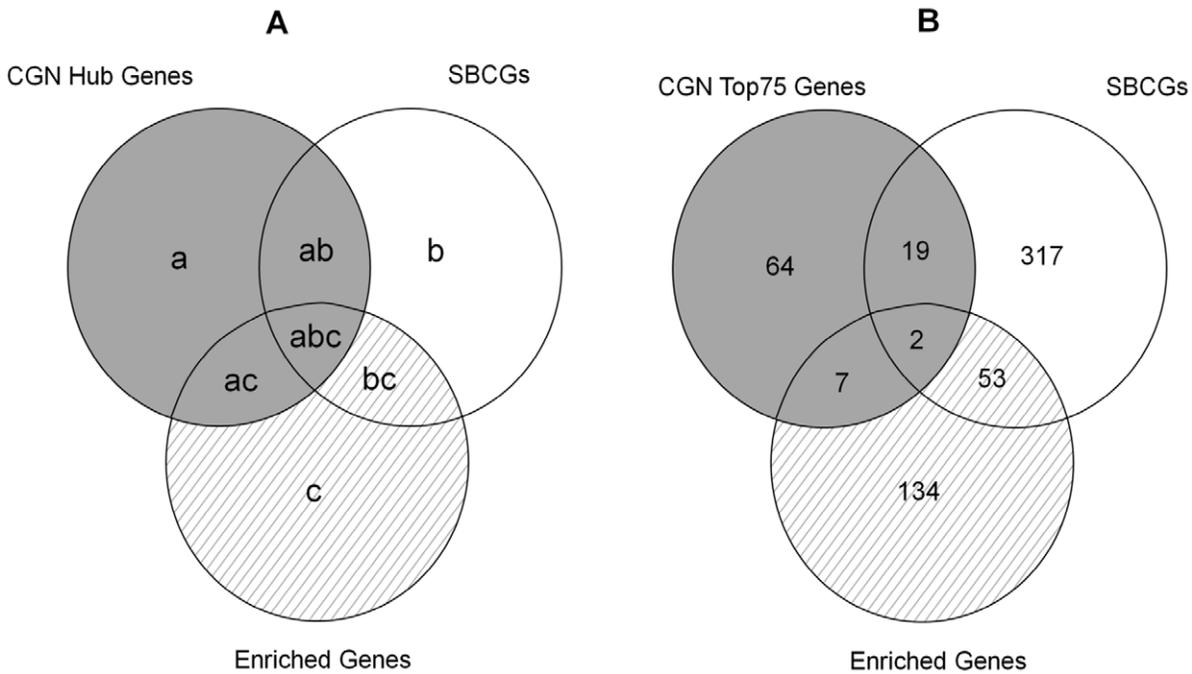
Figure 10. Venn diagrams depicting the strategy used for the identification of SBC-specific targets. A. Venn diagram legend representing generic cancer genes identified by network analysis (CGN hub genes; gray area) and gene subsets used as a source of target genes that are specific to bone metastasis (hatched area). In search of SBC-specific targets, the generic cancer genes were rejected and the potential source of target genes was refined to obtain the final targets. B. Venn diagram depicting components of gene-sets identified using Top75 hubs. The source set of target genes (53+134) was refined to obtain seven targets.
doi:10.1371/journal.pone.0049401.g010

topological correlates of essentiality. We use KEGG-PIC genes as a prototype of generic cancer mechanisms. There is scope for improvement of this data by inclusion of generic genes through

curation from landmark studies [1,2,73,74]. We believe that better metrics that embody structure and information dynamics over the cancer genes interactome could be developed that are more
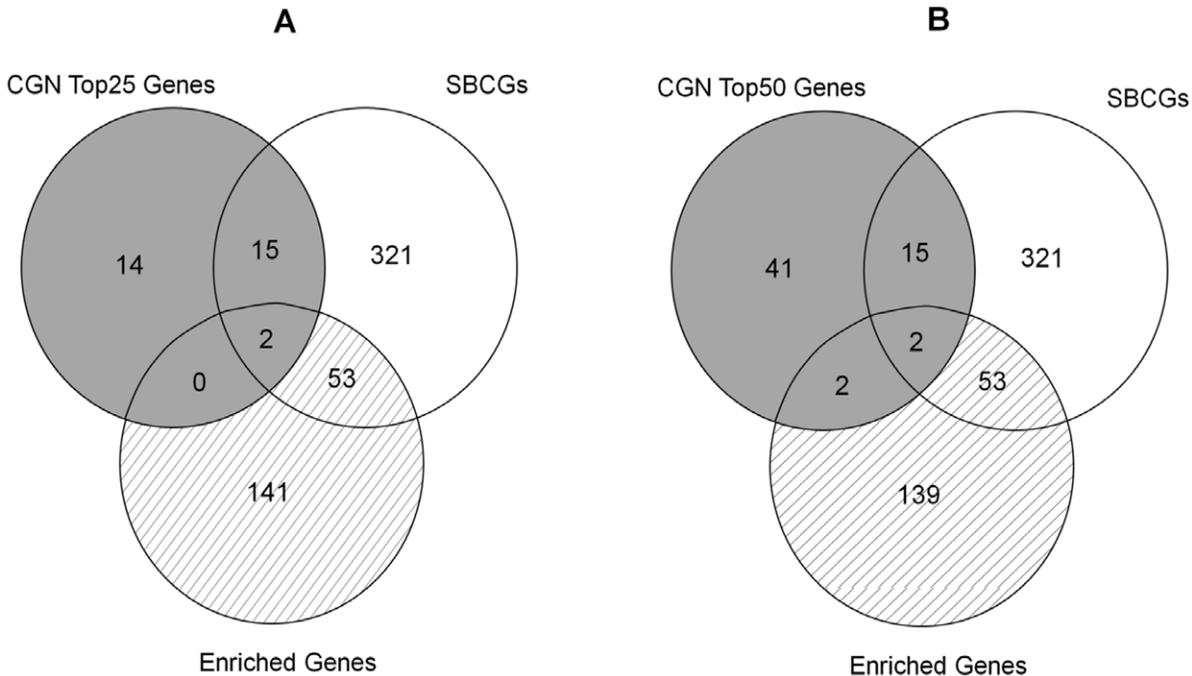

Figure 11. Venn diagram of gene sets obtained with Top25 and Top50 CGN hubs. The Venn diagrams with generic gene-set (gray area) and source set of target genes (hatched area) for (A) Top25 and (B) Top50 CGN hubs.
doi:10.1371/journal.pone.0049401.g011

successful in elucidating the regulatory features characterizing molecular circuitries of cancers.

In view of the complex and subtly intertwined regulatory mechanisms of cancers, we modeled it as a network of protein interactions and aimed to identify generic cancer genes that specify hallmark features of cancers. We find that, indeed hubs of the cancer genes interactome that are central to the structural integrity and dynamical interplay of proteins, correlate with genes from generic pathways of cancers. We believe that the methodology presented here could be useful in obtaining cancer-specific targets of potential therapeutic value.

Materials and Methods

Cancer Genes Network (CGN)

We used HPRD (Human Protein Reference Database) [49], one of the most comprehensive resources of human protein-protein interactions (PPIs), to construct a reference human protein interactome. HPRD is a manually curated human protein-protein interaction resource containing 36617 unique human PPIs and 9427 associated proteins (Release 9: April 13, 2010). It is one of the best resources of human PPIs, containing the largest number of binary non-redundant human PPIs, largest number of genes annotated with at least one interactor, and largest citations of PPIs curated [75].

The data of cancer genes involved in carcinogenesis were compiled from CancerGenes database [48] (as on May 2012). CancerGenes database is a compilation of cancer gene lists annotated by experts with information from key publicly available databases. Cancer associated genes are collected from various sources such as Cancer Map Pathways, Sanger Cancer Gene Census, Sanger Catalogue of Somatic Mutations in Cancer, reviews on cancer, Entrez queries and prostate cancer list. This data of 3164 cancer genes was used to compile a protein interactome representing molecular mechanisms of cancers. The interactions associated with proteins corresponding to these genes were collected from HPRD [49]. This network was called CGN (Cancer Genes Network), in which nodes represent cancer genes and edges represent experimentally validated interactions between a pair of genes i and j. Thus CGN is an intricate network of cancer proteins, comprising of 11602 interactions among 2665 proteins. It contains a giant cluster of 2376 proteins interlinked via 11590 interactions and the rest fragmented into minor clusters and isolated nodes (Figure 3). To adjudge the scale-free nature of degree and betweenness distributions of CGN, we tested the power-law hypothesis and estimated the parameters for these distributions with the technique based on maximum likelihood methods and the Kolmogorov-Smirnov statistic (Figure 4) [52].

Secondary Bone Cancer Genes (SBCGs)

We curated and compiled 391 genes involved in 'metastasis to bone from primary prostate and breast cancer' (167 and 230 genes, respectively), through literature mining (Set-B in Figure 2). These genes were called SBCGs (Secondary Bone Cancer Genes). We used following keywords to search for SBCGs: "secondary bone cancer genes from primary prostate cancer", "secondary bone cancer genes from primary breast cancer" (Pubmed) and "genes involved in metastasis of prostate cancer to bone", "genes involved in metastasis of breast cancer to bone" (Google Scholar). Table S3 provides the details of SBCGs compiled from the literature.

Gene Ontology Analysis

Gene Ontology (GO) enrichment or overrepresentation analysis allows one to identify characteristic biological attributes in a given gene set. It is based on the hypothesis that functionally related genes should accumulate in the corresponding GO category. We used GOrilla [41], a tool to identify enriched GO terms, to obtain biological attributes characterizing CGN hub genes as well as SBCGs. It uses the hypergeometric distribution to identify enriched GO terms in a given set of genes. We performed GO enrichment using 'two unranked lists of genes' mode, with an 'unranked target set' in the background of an 'unranked source set'. We identify 'significantly enriched GO terms', for Biological Process, Molecular Function and Cellular Components, with p-value cut-off of 0.001 and those having at most 100 genes associated with, in the source data (B≤100). The latter criterion is used to weed out terms those are too generic. The CGN Top75 hub genes (target) were enriched against all genes present in CGN (source). Similarly, the SBCGs (target) were enriched against CGN (source). These GO enrichment experiments help us obtain biological attributes that characterize the CGN hubs and those characterizing SBCGs, respectively, in the background of 'cancer genes' universe.

We performed the GO enrichment studies at three different p-values (0.01, 0.001 and 0.0001) for, both, SBCGs and CGN Top75 gene-set. We found that the significantly enriched GO terms obtained for BP, MF and CC were similar for p-values 0.01 and 0.001. For stricter p-value of 0.0001, we found that many of the GO terms relevant for secondary bone cancer and with generic role in cancer mechanisms, respectively, were lost. In the case of SBCGs, 14 BP and 4 MF disease-specific GO terms were lost at p-value 0.0001, including following key GO terms: angiogenesis (GO:0001525), regulation of bone remodelling (GO:0046850), bone mineralization (GO:0030282), collagen fibril organisation (GO:0030199), integrin binding (GO:0005178), laminin binding (GO:0043236), platelet derived growth factor binding (GO:0048407) and regulation of chemotaxis (GO:0050920). In the case of CGN Top75 gene-set analysis, 20 BP and 2 MF GO terms were lost, including following GO terms important for generic cancer mechanisms, at p-value 0.0001: cellular component disassembly involved in apoptotic process (GO:0006921), positive regulation of MAPK cascade (GO:0043410), positive regulation of cell cycle process (GO:0090068), regulation of G1/S transition of mitotic cell cycle (GO:2000045), mitotic cell cycle G1/S transition checkpoint (GO:0031575), growth hormone receptor signaling pathway (GO:0060396), response to fibroblast growth factor stimulus (GO:0071774) and transforming growth factor beta receptor signaling pathway (GO:0007179).

The value of 'B' (number of genes from the source set associated with a given GO term) was chosen to increase the 'signal' (specific GO terms) and to reduce 'noise' (non-specific GO terms). High values of 'B' increase noise by populating the enrichment results with non-specific GO terms; whereas small values of B reduce the signal by rejecting specific and relevant GO terms. We performed multiple experiments at different thresholds of B-values. (i) For B≤50, we missed 6 BP and 3 MF GO terms specific to SBC including the following: positive regulation of cell adhesion, angiogenesis (GO:0001525), cell-cell adhesion (GO:0016337), cytokine activity (GO:0005125), glycosaminoglycan binding (GO:0005539) and carbohydrate binding (GO:0030246). Similarly, in the case of ACGN Top75 gene-set, 12 BP and 3 MF relevant GO terms were lost including DNA damage response, signal transduction by p53 class mediator (GO:0030330), regulation of DNA replication (GO:0006275), positive regulation of cell cycle process (GO:0090068), positive regulation of MAPK cascade

(GO:0043410), Ras protein signal transduction (GO:0007265) and response to UV (GO:0009411). (ii) For 50<B≤100, we got most of the relevant terms for both SBCGs and ACGN Top75 gene-set. (iii) For B>100, almost all the terms obtained were non-specific (noise). Hence, we chose B≤100 to maximize the relevant GO terms and to reduce the non-specific GO terms in the enrichment results.

Protein Interactome Analysis

We performed network analysis of CGN to compute various graph-theoretical metrics, using NetworkAnalyzer plugin of Cytoscape [76]. We computed seven network centrality parameters based on network connectivity (degree [53] and neighborhood connectivity [54]), network flow (betweenness [33,34], stress [33,34], average shortest path length [55]) and local clustering (clustering coefficient [53,55,56] and topological coefficient [57]).

Degree [53] corresponds to the number of nodes adjacent to a given node v, where adjacent means directly connected. The degree distribution $P(k)$ of a network is then defined to be the fraction of nodes in the network with degree k. Thus if there are total N nodes in a network and n_k of them have degree k, we have $P(k) = n_k/N$.

'Neighbourhood connectivity' [54] of a node v is defined as the average connectivity of all neighbors of v. For a node v with k_v number of neighbors, the neighborhood connectivity is defined as:

$$nc(v) = \frac{\sum_{w \in W} k_w}{k_v}$$

Where w is the set of neighbors of v, and k_w is the degree of each of the neighboring node.

'Stress' [33,34] and (its normalized counterpart) 'betweenness' [33,34] enumerate number of shortest paths from all pairs of vertices passing through the node of interest. In a graph $G = (V,E)$ with N nodes, stress ($str(v)$) is defined as the total number of shortest paths passing through a node v:

$$str(v) \equiv \sum_{s \neq v \in N} \sum_{t \neq v \in N} \sigma_{st}(v)$$

Where, $\sigma_{st(v)}$ is the number of shortest paths from s to t that pass through vertex v. Betweenness is normalized (with the total number of shortest paths in graph G; $\sum_{\substack{s \neq v \neq t \in V \\ s \neq t}} \sigma_{st}$) value of stress.

The higher the value of stress/betweenness, the higher is the relevance of the protein as a critical mediator of regulatory molecules and/or functional modules.

The average shortest path length [55] of a vertex v in graph G, corresponds to the average of all the shortest paths between v and the rest of the vertices. The average shortest path length of vertex v is defined as:

$$asp(v) \equiv \sum_{w \neq v \in V} \frac{dist(v,w)}{(N-1)}$$

Where V is the set of nodes in G, $dist(v,w)$ is the shortest path from v to w, and N is the number of nodes in G.

Clustering Coefficient C_v [53,55,56] of a node v is defined as:

$$C_v \equiv \frac{2e_v}{k_v(k_v - 1)}$$

Where k_v is the number of neighbors of v and e_v is the number of connected pairs of nodes between all neighbors of v.

Topological Coefficient $tc(v)$ [57] of a node v is defined as,

$$tc(v) \equiv \frac{\sum_{w \neq v \in V} J(v,w)}{k_v}$$

Where V is the set of nodes in G and $J(v,w)$ is the number of neighbors shared between the nodes v to w, plus one if there is a direct link between v and w.

We used degree, betweenness and stress metrics for identification of hubs of CGN. These metrics have been reported to be useful in identification of hubs of biological relevance [31,35,36] and those relevant to cancer [37].

Identification of Hub Nodes and their Controls

First, we ranked the genes of CGN for each of the chosen parameters (degree, betweenness and stress). Then, we compiled eight 'hub gene-sets' (called Top25, Top50, so on till Top200), containing genes with ranks above the cut-off threshold, for each of the three parameters (Table S1). Each of these hub gene-sets contains hub genes identified by either of the three parameters. Thus defined, the size of a hub gene-set may be up to three times the hub cut-off threshold, depending upon the similarity between the hubs identified by each of the parameters. Top25, Top50 and Top75 hub gene-sets of CGN contain 31, 60 and 92 hub genes, respectively. Figure 3 depicts the CGN hub genes identified using Top75 hubs criterion. As a negative control for these hub gene-sets, we identified the corresponding genes (from the bottom of ranked lists), with lowest ranking for the chosen parameters. As random controls, we randomly sample a corresponding number of genes from CGN (1000 instances each).

Compilation of Genes Involved in Generic Cancer Mechanisms (KEGG-PIC)

Cancer cell mechanisms could be envisaged as an elaborate integrated circuit of intracellular signaling networks [1,2]. This map of molecular mechanisms could be represented as a combination of circuits and subcircuits with considerable cross talk among them [1]. We compiled a set of (328) representative genes known to be implicated in generic cancer mechanisms from cancer pathways/circuits (Table S2). 'Pathways in Cancer' (PIC) (hsa05200) from KEGG PATHWAY database, are representative of generic cancer circuits. Here, we call this 'generic cancer genes' set KEGG-PIC [62]. KEGG-PIC comprise the following KEGG pathways: colorectal cancer (hsa05210), pancreatic cancer (hsa05212), thyroid cancer (hsa05216), acute myeloid leukemia (hsa05221), chronic myeloid leukemia (hsa05220), basal cell carcinoma (hsa05217), melanoma (hsa05218), renal cell carcinoma (hsa05211), bladder cancer (hsa05219), prostate cancer (hsa05215), endometrial cancer (hsa05213), small cell lung cancer (hsa05222), non-small cell lung cancer (hsa05223) and glioma (hsa05214).

Mouse Phenotype Data

For inferring the biological significance of the network parameters, we divided CGN into two sets, essential and non-essential, using the phenotypic information of the corresponding mouse ortholog [47,77]. It is assumed that human orthologs of mouse genes could be mapped onto each other for their function and biological essentiality. We considered the classes of embryonic, perinatal, neonatal or postnatal lethality in mouse models as lethal phenotypes, and the rest of the phenotypes as non-lethal ones. The human orthologs of murine genes were considered as essential, when the murine gene was annotated with one of the following phenotypes [63]: neonatal lethality (MP:0002058), embryonic lethality (MP:0002080), perinatal lethality (MP:0002081), postnatal lethality (MP:0002082), lethality-postnatal (MP:0005373), lethality-embryonic/perinatal (MP:0005374), embryonic lethality before implantation (MP:0006204), embryonic lethality before somite formation (MP:0006205) or embryonic lethality before turning of embryo (MP:0006206). The human-mouse orthology and mouse phenotype data was obtained from Mouse Genome Informatics [65] (May 2012). Out of a total 2665 cancer genes of CGN, 1315 (49.34%) were essential genes and the rest 1350 (50.66%) were non-essential. For each hub definition, from the hub-genes identified using each of the seven network metric, we compute the percentage of hubs that are essential (Figure 8A). We also found the percentage of essential genes in the corresponding non-hubs (Figure 8B).

Venn Diagram

We logically juxtaposed the hubs of cancer genes network, curated bone cancer metastasis genes and cancer genes annotated with characteristic SBC GO terms to identify SBC-specific targets (Figure 2). Figure 10A illustrates this process highlighting the hubs of CGN (shaded area) that correspond to generic cancer genes and subsets potentially containing genes specific to SBC (hatched area). Figure 10B depicts the data when Top75 hubs of CGN are taken into consideration. For this data, there are 92 hubs of which 21 (Set-ab and Set-abc) happen to be common with SBCGs and 9 of those (Set-ac and Set-abc) are common to 'enriched genes'. Among the 'secondary bone cancer enriched cancer genes' 55 are common to SBCGs (Set-bc and Set-abc). We find that, for this data, there are 2 genes (Set-abc) that happen to be CGN hubs that are common to SBCGs and are also annotated with characteristic SBC GO terms. The logical juxtaposition results for the data of Top25 and Top50 hubs of CGN are depicted in Figure 11A and Figure 11B, respectively.

Prediction of SBC-specific Candidate Genes

We propose that the genes that are specific to secondary bone cancer mechanisms would be annotated with characteristic GO terms that are obtained from overrepresentation analysis of a literature curated list of genes implicated in metastasis to bone. From the 'secondary bone cancer enriched cancer genes', that serve as a 'source set of targets' we filtered the CGN hubs (Set-ac and Set-abc; shaded area) as they are generic to cancers (Figure 6) and found to be correlating with essential genes (Figure 8). Towards our aim of identifying SBC-specific targets, we focused on genes in Set-c and Set-bc (hatched area in Figure 10). We identified SBC-specific targets by refining these sets of genes to obtain genes that are annotated with any of the GO terms representing 'bone processes' as well as that of 'metastasis' (Figure 9) (Table S4). .

Supporting Information

In the supporting information we present the details of hubs of CGN; the KEGG-PIC genes that serve as a reference set of generic cancer genes; secondary bone cancer genes that were curated and compiled; characteristic GO terms that were used to obtain SBC-specific targets and relevance of SBC-specific targets identified from experimentally validated studies. The supporting information contains 5 tables, out of which Table S3 has 46 references, Table S4 has 23 references and Table S5 has 8 references.

Acknowledgements

Supporting Information

Table S1 Hub genes (Top25–Top200) of Cancer Genes Network.
(PDF)

Table S2 Details of KEGG-PIC genes (328) from KEGG PATHWAY Database.
(PDF)

Table S3 SBCGs (391) compiled for metastasis of primary breast and prostate cancer.
(PDF)

Table S4 Significantly enriched GO terms, characteristic to metastasis to bone, identified from enrichment analysis of SBCGs.
(PDF)

Table S5 Relevance of SBC targets.
(PDF)

Author Contributions

Conceived and designed the experiments: SV GB. Performed the experiments: SV. Analyzed the data: SV GB. Wrote the paper: SV GB.

References

1. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144: 646–674. doi:10.1016/j.cell.2011.02.013.
2. Hanahan D, Weinberg RA, Francisco S (2000) The Hallmarks of Cancer. Cell 100: 57–70. doi:10.1016/S0092–8674(00)81683–9.
3. Barabási A-L, Gulbahce N, Loscalzo J (2011) Network medicine: a network-based approach to human disease. Nature reviews Genetics 12: 56–68. doi:10.1038/nrg2918.
4. Ergün A, Lawrence CA, Kohanski MA, Brennan TA, Collins JJ (2007) A network biology approach to prostate cancer. Molecular systems biology 3: 82. doi:10.1038/msb4100125.
5. Chuang H-Y, Lee E, Liu Y-T, Lee D, Ideker T (2007) Network-based classification of breast cancer metastasis. Molecular systems biology 3: 140. doi:10.1038/msb4100180.
6. Chang W, Ma L, Lin L, Gu L, Liu X, et al. (2009) Identification of novel hub genes associated with liver metastasis of gastric cancer. International journal of cancer Journal international du cancer 125: 2844–2853. doi:10.1002/ijc.24699.
7. Talmadge JE, Fidler IJ (2010) The biology of cancer metastasis: historical perspective. Cancer research 70: 5649–5669. doi:10.1158/0008–5472.CAN-10–1040.

8. Chaffer CL, Weinberg RA (2011) A Perspective on Cancer Cell Metastasis. Science 331: 1559–1564. doi:10.1126/science.1203543.

9. Virk MS, Lieberman JR (2007) Tumor metastasis to bone. Arthritis research & therapy 9: S5. doi:10.1186/ar2169.

10. Chu K, Cheng C-J, Ye X, Lee Y-C, Zurita AJ, et al. (2008) Cadherin-11 promotes the metastasis of prostate cancer cells to bone. Molecular Cancer Research 6: 1259–1267. doi:10.1158/1541-7786.MCR-08-0077.

11. Kozlow W, Guise TA (2005) Breast cancer metastasis to bone: mechanisms of osteolysis and implications for therapy. Journal of Mammary Gland Biology and Neoplasia 10: 169–180. doi:10.1007/s10911-005-5399-8.

12. Coleman RE (1997) Skeletal complications of malignancy. Cancer 80: 1588–1594. doi:10.1002/(SICI)1097–0142(19971015)80:8+<1588::AID–CNCR9>3.0.CO;2–G..

13. Mundy GR (2002) Metastasis to bone: causes, consequences and therapeutic opportunities. Nature Reviews Cancer 2: 584–593. doi:10.1038/nrc867.

14. Hess KR, Varadhachary GR, Taylor SH, Wei W, Raber MN, et al. (2006) Metastatic patterns in adenocarcinoma. Cancer 106: 1624–1633. doi:10.1002/cncr.21778.

15. Bu G, Lu W, Liu C-C, Selander K, Yoneda T, et al. (2008) Breast cancer-derived Dickkopf1 inhibits osteoblast differentiation and osteoprotegerin expression: implication for breast cancer osteolytic bone metastases. International Journal of Cancer 123: 1034–1042. doi:10.1002/ijc.23625.

16. Cicek M, Oursler MJ (2006) Breast cancer bone metastasis and current small therapeutics. Cancer Metastasis Reviews 25: 635–644. doi:10.1007/s10555-006-9035-x.

17. Guise TA (2009) Breaking down bone: new insight into site-specific mechanisms of breast cancer osteolysis mediated by metalloproteinases. Genes & Development 23: 2117–2123. doi:10.1101/gad.1854909.

18. Lu X, Wang Q, Hu G, Van Poznak C, Fleisher M, et al. (2009) ADAMTS1 and MMP1 proteolytically engage EGF-like ligands in an osteolytic signaling cascade for bone metastasis. Genes & Development 23: 1882–1894. doi:10.1101/gad.1824809.

19. Smid M, Wang Y, Klijn JGM, Sieuwerts AM, Zhang Y, et al. (2006) Genes associated with breast cancer metastatic to bone. Journal of Clinical Oncology 24: 2261–2267. doi:10.1200/JCO.2005.03.8802.

20. Weigelt B, Peterse J (2005) Breast cancer metastasis: markers and models. Nature reviews cancer 5: 591–602. doi:10.1038/nrc1670.

21. Zhang XH-F, Wang Q, Gerald W, Hudis CA, Norton L, et al. (2009) Latent bone metastasis in breast cancer tied to Src-dependent survival signals. Cancer Cell 16: 67–78. doi:10.1016/j.ccr.2009.05.017.

22. Sloan EK, Anderson RL (2002) Genes involved in breast cancer metastasis to bone. Cellular and Molecular Life Sciences 59: 1491–1502. doi:10.1007/s00018-002-8524-5.

23. Zhang X, Wang W, True LD, Vessella RL, Takayama TK (2009) Protease-activated receptor-1 is upregulated in reactive stroma of primary prostate cancer and bone metastasis. The Prostate 69: 727–736. doi:10.1002/pros.20920.

24. Taichman RS, Cooper C, Keller ET, Pienta KJ, Taichman NS, et al. (2002) Use of the Stromal Cell-derived Factor-1/CXCR4 Pathway in Prostate Cancer Metastasis to Bone. Cancer Research 62: 1832–1837.

25. Valta MP, Tuomela J, Bjartell A, Valve E, Väänänen HK, et al. (2008) FGF-8 is involved in bone metastasis of prostate cancer. International Journal of Cancer 123: 22–31. doi:10.1002/ijc.23422.

26. Secondini C, Wetterwald A, Schwaninger R, Thalmann GN, Cecchini MG (2011) The role of the BMP signaling antagonist noggin in the development of prostate cancer osteolytic bone metastasis. PLoS One 6: e16078. doi:10.1371/journal.pone.0016078.

27. Koreckij T, Nguyen H, Brown LG, Yu EY, Vessella RL, et al. (2009) Dasatinib inhibits the growth of prostate cancer in bone and provides additional protection from osteolysis. British Journal of Cancer 101: 263–268. doi:10.1038/sj.bjc.6605178.

28. Bailey CL, Kelly P, Casey PJ (2009) Activation of Rap1 promotes prostate cancer metastasis. Cancer Research 69: 4962–4968. doi:10.1158/0008–5472.CAN-08–4269.

29. Chen G, Sircar K, Aprikian A, Potti A, Goltzman D, et al. (2006) Expression of RANKL/RANK/OPG in primary and metastatic human prostate cancer as markers of disease stage and functional regulation. Cancer 107: 289–298. doi:10.1002/cncr.21978.

30. Jones DH, Nakashima T, Sanchez OH, Kozieradzki I, Komarova SV, et al. (2006) Regulation of cancer cell migration and bone metastasis by RANKL. Nature 440: 692–696. doi:10.1038/nature04524.

31. Jeong H, Mason SP, Barabási AL, Oltvai ZN (2001) Lethality and centrality in protein networks. Nature 411: 41–42. doi:10.1038/35075138.

32. Batada NN, Hurst LD, Tyers M (2006) Evolutionary and physiological importance of hub proteins. PLoS Computational Biology 2: e88. doi:10.1371/journal.pcbi.0020088.

33. Freeman L (1977) A Set of Measures of Centrality Based on Betweenness. Sociometry 40: 35–41.

34. Brandes U (2001) A faster algorithm for betweenness centrality. Journal of Mathematical Sociology 25: 163–177. doi:10.1080/0022250X.2001.9990249.

35. Joy MP, Brock A, Ingber DE, Huang S (2005) High-betweenness proteins in the yeast protein interaction network. Journal of Biomedicine & Biotechnology 2005: 96–103. doi:10.1155/JBB.2005.96.

36. Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M (2007) The importance of bottlenecks in protein networks: correlation with gene essentiality and expression

37. Jonsson PF, Bates PA (2006) Global topological features of cancer proteins in the human interactome. Bioinformatics 22: 2291–2297. doi:10.1093/bioinformatics/btl390.

38. Gene T, Consortium O (2001) Creating the Gene Ontology Resource□: Design and Implementation. Genome Research 11: 1425–1433. doi:10.1101/gr.180801.examining.

39. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide. PNAS 102: 15545–15550. doi:10.1073/pnas.0506580102.

40. Rhee SY, Wood V, Dolinski K, Draghici S (2008) Use and misuse of the gene ontology annotations. Nature reviews Genetics 9: 509–515. doi:10.1038/nrg2363.

41. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10: 48. doi:10.1186/1471–2105–10–48.

42. Huang DW, Sherman BT, Lempicki R a (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic acids research 37: 1–13. doi:10.1093/nar/gkn923.

43. Emmert-Streib F (2007) The chronic fatigue syndrome: a comparative pathway analysis. Journal of computational biology 14: 961–972. doi:10.1089/cmb.2007.0041.

44. Jia P, Kao C-F, Kuo P-H, Zhao Z (2011) A comprehensive network and pathway analysis of candidate genes in major depressive disorder. BMC Systems Biology 5: S12. doi:10.1186/1752-0509-5-S3-S12.

45. Cerami E, Demir E, Schultz N, Taylor BS, Sander C (2010) Automated network analysis identifies core pathways in glioblastoma. PloS one 5: e8918. doi:10.1371/journal.pone.0008918.

46. Lascorz J, Hemminki K, Försti A (2011) Systematic enrichment analysis of gene expression profiling studies identifies consensus pathways implicated in colorectal cancer development. Journal of carcinogenesis 10: 7. doi:10.4103/1477–3163.78268.

47. Ortutay C, Vihinen M (2009) Identification of candidate disease genes by integrating Gene Ontologies and protein-interaction networks: case study of primary immunodeficiencies. Nucleic Acids Research 37: 622–628. doi:10.1093/nar/gkn982.

48. Higgins ME, Claremont M, Major JE, Sander C, Lash AE (2007) CancerGenes: a gene selection resource for cancer genome projects. Nucleic Acids Research 35: D721–6. doi:10.1093/nar/gkl811.

49. Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, et al. (2003) Development of Human Protein Reference Database as an Initial Platform for Approaching Systems Biology in Humans. Genome Research 13: 2363–2371. doi:10.1101/gr.1680803.17.

50. Dorogovtsev S (2004) The shortest path to complex networks. Arxiv preprint cond-mat/0404593: 1–25.

51. Albert R, Barabasi A-L (2002) Statistical mechanics of complex networks. Reviews of Modern Physics 74. doi:10.1103/RevModPhys.74.47.

52. Clauset A, Shalizi CR, Newman MEJ (2009) Power-Law Distributions in Empirical Data. SIAM Review 51: 661–703. doi:10.1137/070710111.

53. Barabási A-L, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nature reviews Genetics 5: 101–113. doi:10.1038/nrg1272.

54. Maslov S, Sneppen K (2002) Specificity and stability in topology of protein networks. Science 296: 910–913. doi:10.1126/science.1065103.

55. Watts DJ, Strogatz SH (1998) Collective dynamics of `small-world' networks. Nature 393: 440–442. doi:10.1038/30918.

56. Newman MEJ, Watts DJ, Strogatz SH (2002) Random graph models of social networks. Proceedings of the National Academy of Sciences of the United States of America 99 Suppl 1: 2566–2572. doi:10.1073/pnas.012582999.

57. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, et al. (2005) A human protein-protein interaction network: a resource for annotating the proteome. Cell 122: 957–968. doi:10.1016/j.cell.2005.08.029.

58. Wyllie AH, Kerr JFR, Currie AR (1980) Cell death: the significance of apoptosis. International Review of Cytology 68: 251–306. doi:10.1016/S0074-7696(08)62312-8.

59. Harris CC (1996) P53 Tumor Suppressor Gene: From the Basic Research Laboratory To the Clinic–an Abridged Historical Perspective. Carcinogenesis 17: 1187–1198. doi:10.1093/carcin/17.6.1187.

60. Levine AJ (1997) P53, the Cellular Gatekeeper for Growth and Division. Cell 88: 323–331. doi:10.1016/S0092–8674(00)81871-1.

61. Dai JL, Bansal RK, Kern SE (1999) G1 cell cycle arrest and apoptosis induction by nuclear Smad4/Dpc4: phenotypes reversed by a tumorigenic mutation. Proceedings of the National Academy of Sciences of the United States of America 96: 1427–1432. doi:10.1073/pnas.96.4.1427.

62. KEGG-Pathways in cancer: hsa05200 (n.d.). Available:http://www.genome.jp/dbget-bin/www_bget?pathway+hsa05200.

63. Smith CL, Goldsmith C-AW, Eppig JT (2005) The Mammalian Phenotype Ontology as a tool for annotating, analyzing and comparing phenotypic information. Genome Biology 6: R7. doi:10.1186/gb-2004-6-1-r7.

64. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, et al. (2007) The human disease network. Proceedings of the National Academy of Sciences 104: 8685–8690. doi:10.1073/pnas.0701361104.

65. Blake JA, Bult CJ, Kadin JA, Richardson JE, Eppig JT (2011) The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics. Nucleic Acids Research 39: D842-8. doi:10.1093/nar/gkq1008.

66. Rangaswami H, Bulbule A, Kundu GC (2006) Osteopontin: role in cell signaling and cancer progression. Trends in cell biology 16: 79–87. doi:10.1016/j.tcb.2005.12.005.

67. Rittling SR, Chambers AF (2004) Role of osteopontin in tumour progression. British journal of cancer 90: 1877–1881. doi:10.1038/sj.bjc.6601839.

68. Wang H-S, Hung Y, Su C-H, Peng S-T, Guo Y-J, et al. (2005) CD44 cross-linking induces integrin-mediated adhesion and transendothelial migration in breast cancer cell line by up-regulation of LFA-1 (alpha L beta2) and VLA-4 (alpha4beta1). Experimental cell research 304: 116–126. doi:10.1016/j.yexcr.2004.10.015.

69. Kang Y, Siegel PM, Shu W, Drobnjak M, Kakonen SM, et al. (2003) A multigenic program mediating breast cancer metastasis to bone. Cancer cell 3: 537–549.

70. Mishina Y, Starbuck MW, Gentile MA, Fukuda T, Kasparcova V, et al. (2004) Bone morphogenetic protein type IA receptor signaling regulates postnatal osteoblast function and bone remodeling. The Journal of biological chemistry 279: 27560–27566. doi:10.1074/jbc.M404222200.

71. Matsumoto K, Takayama N, Ohnishi J, Ohnishi E, Shirayoshi Y, et al. (2001) Tumour invasion and metastasis are promoted in mice deficient in tenascin-X. Genes to cells 6: 1101–1111.

72. Eppert K, Wunder JS, Aneliunas V, Kandel R, Andrulis IL (2005) von Willebrand factor expression in osteosarcoma metastasis. Modern pathology 18: 388–397. doi:10.1038/modpathol.3800265.

73. Hahn WC, Weinberg RA (2002) Modelling the molecular circuitry of cancer. Nature reviews Cancer 2: 331–341. doi:10.1038/nrc795.

74. Vogelstein B, Kinzler KW (2004) Cancer genes and the pathways they control. Nature medicine 10: 789–799. doi:10.1038/nm1087.

75. Mathivanan S, Periaswamy B, Gandhi TKB, Kandasamy K, Suresh S, et al. (2006) An evaluation of human protein-protein interaction data in the public domain. BMC bioinformatics 7: S19. doi:10.1186/1471–2105–7-S5-S19.

76. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research 13: 2498–2504. doi:10.1101/gr.1239303.

77. Yildirim MA, Goh K-I, Cusick ME, Barabási A-L, Vidal M (2007) Drug-target network. Nature biotechnology 25: 1119–1126. doi:10.1038/nbt1338.