

RESEARCH ARTICLE

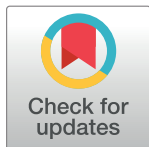
Graph-theoretical model of global human interactome reveals enhanced long-range communicability in cancer networks

Evgeny Gladilin^{1,2✉*}

1 Division of Theoretical Bioinformatics, German Cancer Research Center, Berliner Str. 41, 69120 Heidelberg, Germany, **2** BioQuant and IPMB, University Heidelberg, Im Neuenheimer Feld 267, 69120 Heidelberg, Germany

✉ Current address: Leibniz Institute of Plant Genetics and Crop Plant Research, Corrensstrasse 3, 06466 Gatersleben, Germany

* evgeny.gladilin@gmail.com



Abstract

Malignant transformation is known to involve substantial rearrangement of the molecular genetic landscape of the cell. A common approach to analysis of these alterations is a reductionist one and consists of finding a compact set of differentially expressed genes or associated signaling pathways. However, due to intrinsic tumor heterogeneity and tissue specificity, biomarkers defined by a small number of genes/pathways exhibit substantial variability. As an alternative to compact differential signatures, global features of genetic cell machinery are conceivable. Global network descriptors suggested in previous works are, however, known to potentially be biased by overrepresentation of interactions between frequently studied genes-proteins. Here, we construct a cellular network of 74538 directional and differential gene expression weighted protein-protein and gene regulatory interactions, and perform graph-theoretical analysis of global human interactome using a novel, degree-independent feature—the normalized total communicability (NTC). We apply this framework to assess differences in total information flow between different cancer (BRCA/COAD/GBM) and non-cancer interactomes. Our experimental results reveal that different cancer interactomes are characterized by significant enhancement of long-range NTC, which arises from circulation of information flow within robustly organized gene subnetworks. Although enhancement of NTC emerges in different cancer types from different genomic profiles, we identified a subset of 90 common genes that are related to elevated NTC in all studied tumors. Our ontological analysis shows that these genes are associated with enhanced cell division, DNA replication, stress response, and other cellular functions and processes typically upregulated in cancer. We conclude that enhancement of long-range NTC manifested in the correlated activity of genes whose tight coordination is required for survival and proliferation of all tumor cells, and, thus, can be seen as a graph-theoretical equivalent to some hallmarks of cancer. The computational framework for differential network analysis presented herein is of potential interest for a wide range of network perturbation problems given by single or multiple gene-protein activation-inhibition.

OPEN ACCESS

Citation: Gladilin E (2017) Graph-theoretical model of global human interactome reveals enhanced long-range communicability in cancer networks. PLoS ONE 12(1): e0170953. doi:10.1371/journal.pone.0170953

Editor: Peter Csermely, Semmelweis University, HUNGARY

Received: February 24, 2016

Accepted: January 13, 2017

Published: January 31, 2017

Copyright: © 2017 Evgeny Gladilin. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The author received no specific funding for this work.

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

Introduction

Clinically relevant, macroscopically detectable tumors are known to exhibit phenotypic and molecular genetic heterogeneity [1]. Despite considerable genetic diversity, different tumor cells manage to maintain common functional capabilities that manifest in hallmarks of cancer [2]. The underlying mechanisms of cancer hallmark maintenance in different tumors with different genomic profiles are not yet well understood. As a consequence of cancer heterogeneity and plasticity, differential signatures defined by a relatively small number of genes-proteins exhibit substantial variability, which complicates the identification of cancer-specific alterations in microarrays and other omics data.

An alternative approach to quantitative characterization of malignant transformations consists in the assessment of the global architecture of cellular networks. Recent advances in network science provide a powerful theoretical framework for the description of global properties of physical, social and biological networks [3–5]. For construction of binary and weighted biological networks, gene co-expression maps [6–8], pairwise physical interactions and non-physical associations between proteins, DNA, RNA, metabolites and gene regulatory events have been applied [9–23]. Diverse parameters of local and global network organization have been used for quantitative description and differentiation of normal, diseased and random interactomes including graph-theoretical measures such as node degree, centrality, modularity, clustering, [24–27], network statistics [28], information content [29] and hyperbolicity [30]. Global information-theoretical features, such as network entropy, have been shown to significantly differ between cancer and non-cancer interactomes [31, 32].

Cancer networks have repeatedly been reported to be significantly larger, interlinked more densely and more tautly organized in comparison to non-cancer and, in particular, random networks [25, 33–37]. These findings were, however, challenged by reasonable criticism that refers to potential biases of existing network descriptors due to overrepresentation of disease-related genes. Consequently, these genes exhibit a higher number of interactions, higher degrees and other artificially exceptional features in contrast to poorly studied targets [38, 39]. To overcome shortcomings of degree-based descriptors, we present a novel degree-normalized communicability measure that is applied to study information flow in global cancer and non-cancer networks whose basic topology is defined by directional and gene expression weighted protein-protein and gene regulatory interactions.

The manuscript is organized as follows. First, methods for construction of gene expression weighted network topology are described. The experimental results of comparative analysis of cancer and non-cancer interactomes are presented and discussed. The complete set of raw and processed data used in this work can be found in supplementary information.

Methods

Microarray data preprocessing

TCGA level-3 microarray data from tumor and normal tissue samples of breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD) and glioblastoma (GBM) patients are used. Lists of all TCGA samples used in this study are in [S1 Table](#).

Statistical significance of differential gene expression between tumor and normal tissue samples is evaluated using the t-test with the p-value threshold $p < 0.01$. For significantly up/downregulated genes, the log₂-fold average differential gene expression (ADGE) Δ_i is computed. Unidentified and non-significantly altered genes are assumed to have unchanged level of expression ($\Delta_i = 0$). Next, all N genes are sorted according to a hybrid score based on a product of t- and ADGE-values: $\lambda_i = \text{sign}(\Delta_i)(t_i \Delta_i)$. To avoid dependency of subsequent calculations

on statistical outliers, absolute values of gene scores are subsequently substituted by a uniform pattern of average gene expression ranging in $\lambda_i \in [-6.5, 6.5]$. This transformation has the effect that genes with the same rank in a λ_i -sorted list become equal weights in different cancer and non-cancer samples:

$$\begin{aligned} \lambda_i^{\text{brca}} &= \lambda_i^{\text{coad}} = \lambda_i^{\text{gbm}} = \lambda_i^{\text{rand}}, \\ \lambda_1 &> \lambda_2 > \dots > 0 \dots > \lambda_{N-1} > \lambda_N, \\ \lambda_1 &= \max(\lambda_i) = -\min(\lambda_i) = -\lambda_N. \end{aligned} \tag{1}$$

Sorted lists of rank-normalized gene weights for all tumor/norm, norm/tumor (i.e., reversely weighted tumor/norm lists) as well as randomized data are in [S2 Table](#).

Network topology compilation

Network topology is compiled on the basis of directed pairwise protein-protein and gene regulatory interactions by integration of open-source data provided with STRING ([string-db.org](#)), MSigDB ([software.broadinstitute.org/gsea/msigdb](#)) and PATHWAYCOMMONS ([www.pathwaycommons.org](#)). The complete list of 74538 directed pairwise interactions is in [S3 Table](#).

Network communicability: plausibility considerations

First, we want to define a plausible measure for quantification of total information flow (communicability) along a single linear pathway. This should be done in such a way that the absence or malfunction of one single pathway link (i.e., network edge) results in interruption or significant impairment of the entire pathway communicability, see [Fig 1\(a\)](#). This intuitively comprehensible constraint is considered by a measure that is defined as a product of all edge weights E_i , i.e.,

$$\prod_{i=1}^N E_i, \tag{2}$$

where $E_i \geq 0$ are positive numbers whose values indicate working (conducting) or non-working (non-conducting) state of the i -th pathway link. In the case of unweighted networks, conducting and non-conducting states of pathway links are described by binary weights of network edges $E_i = 1$ and $E_i = 0$. In weighted networks, weights of network edges are positive floating-point numbers $E_i > 0$.

If two network nodes are connected by multiple pathways, total information flow should not critically depend on the state of a single pathway link or even one single pathway, see [Fig 1\(b\)](#). Consequently, total communicability between each two network nodes can be defined by a sum of all single pathway communicabilities:

$$\sum_{j=1}^M \left(\prod_{i=1}^{N_j} E_i \right). \tag{3}$$

Another plausible requirement on the network communicability measure is that intensity of information flow should decline with increasing distance from the source. Consequently,

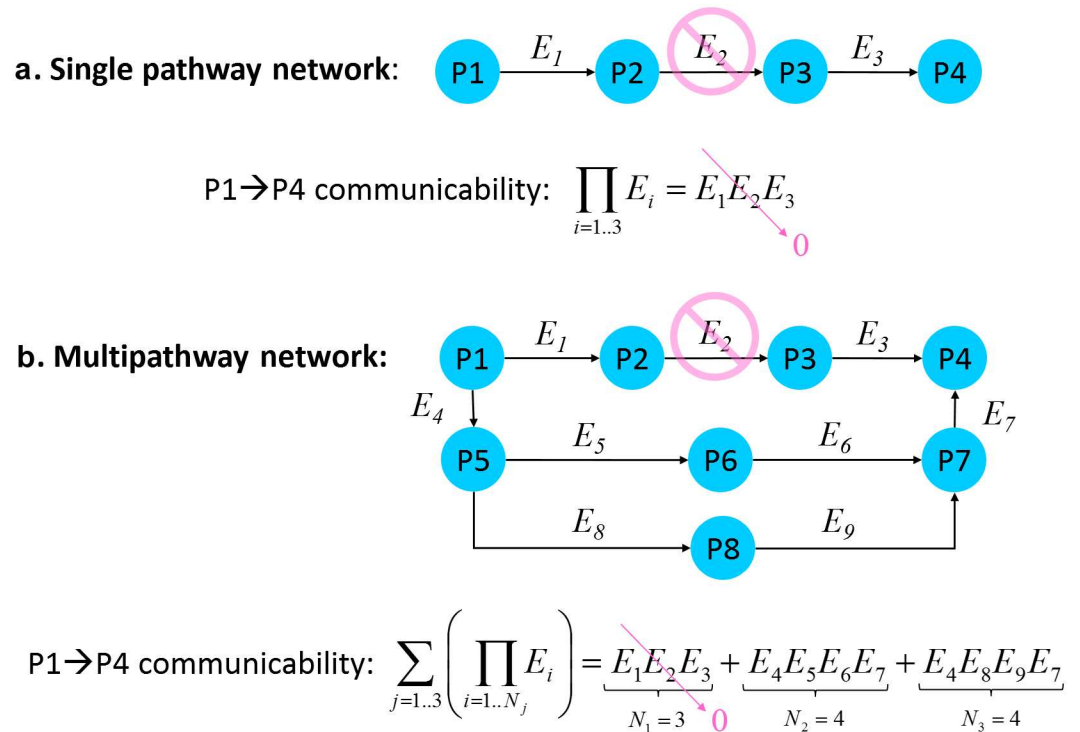


Fig 1. Principle concept of network communicability assessment. **a.** Information flow along a single linear pathway joining two nodes (P1, P4) is defined as a product of weights E_i of all pathway links (here: $E_1 E_2 E_3$). Disruption of a single pathway link ($E_2 \rightarrow 0$) results in interruption of the entire pathway communicability. **b.** Total information flow through multiple pathways is defined as a sum of communicabilities of all linear pathways joining a pair of nodes (here: three pathways with communicabilities $E_1 E_2 E_3$, $E_4 E_5 E_6 E_7$, $E_4 E_8 E_9 E_7$). Disruption of a single pathway link ($E_2 \rightarrow 0$) does not interrupt the total $P1 \rightarrow P4$ communicability.

doi:10.1371/journal.pone.0170953.g001

Eq (3) can be extended to

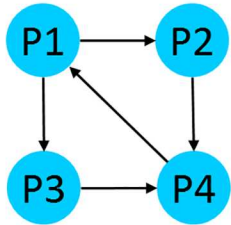
$$\sum_{j=1}^M \left(\frac{1}{\omega(N_j)} \prod_{i=1}^{N_j} E_i \right), \tag{4}$$

where $\omega(N_j) > 1$ denotes a pathway length dependent weighting factor.

Normalized total communicability

In graph theory, the total number of walks of the length n joining nodes of an arbitrary complex network is calculated as the n -power A^n of the graph-representing, sparse adjacency matrix A , see Fig 2. In fact, one can show that Eq (3) is formally identical to the A^n compilation rule from the entries of A . In turn, the weighted version of our plausibly derived communicability measure (Eq (4)) naturally emerges within the concept of the adjacency matrix exponential e^A . Following [40, 41], the total communicability $C_{ij}(n)$ between a pair of network nodes (i, j) joined by all possible walks of the maximum length n is calculated as the exponential of

Directed graph:



All pathways of length n=1:

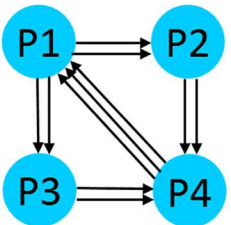
- (1) P1 → P2; A(1,2)=1; W(1,2)= 0.7;
- (1) P1 → P3; A(1,3)=1; W(1,3)= 1.2;
- (1) P2 → P4; A(2,4)=1; W(2,4)= 0.7;
- (1) P3 → P4; A(3,4)=1; W(3,4)= 1.1;
- (1) P4 → P1; A(4,1)=1; W(4,1)= 1.1;

Adjacency matrices (A,W):

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} \quad W = \begin{pmatrix} 0 & 0.7 & 0.7 & 0 \\ 0 & 0 & 0 & 1.2 \\ 0 & 0 & 0 & 1.1 \\ 1.1 & 0 & 0 & 0 \end{pmatrix}$$

unweighted walks weighted walks

All pathways of length n=2:

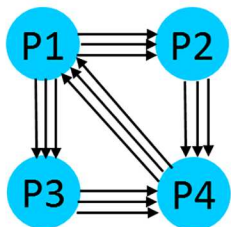


- (2) P1 → P4: P1 → P2 → P4; P1 → P3 → P4;
- (1) P2 → P1: P2 → P4 → P1;
- (1) P3 → P1: P3 → P4 → P1;
- (1) P4 → P3: P4 → P1 → P3;
- (1) P4 → P2: P4 → P1 → P2;

$$A^2 = \begin{pmatrix} 0 & 0 & 0 & 2 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix} \quad W^2 = \begin{pmatrix} 0 & 0 & 0 & 1.6 \\ 1.3 & 0 & 0 & 0 \\ 1.2 & 0 & 0 & 0 \\ 0 & 0.8 & 0.8 & 0 \end{pmatrix}$$

unweighted walks weighted walks

All pathways of length n=3:



- (2) P1 → P1: P1 → P2 → P4 → P1; P1 → P3 → P4 → P1;
- (1) P2 → P2: P2 → P4 → P1 → P2;
- (1) P3 → P3: P3 → P4 → P1 → P3;
- (2) P4 → P4: P4 → P1 → P3 → P4; P4 → P1 → P2 → P4;

$$A^3 = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix} \quad W^3 = \begin{pmatrix} 1.8 & 0 & 0 & 0 \\ 0 & 0.9 & 0.9 & 0 \\ 0 & 0.8 & 0.8 & 0 \\ 0 & 0 & 0 & 1.8 \end{pmatrix}$$

unweighted walks weighted walks

Fig 2. Example of network communicability calculus using adjacency matrices. Given initial adjacency matrices of directional weighted (W) and unweighted (A) pairwise interactions between network nodes, all weighted and unweighted walks of the length $n > 1$ are calculated from the n -power matrices W^n and A^n , respectively. Note that the shortest possible loop in a directed graph has the length $n = 3$. Consequently, diagonal entries of W^3 and A^3 contain non-zero values, while the $n = 1, 2$ power matrices are hollow.

doi:10.1371/journal.pone.0170953.g002

the adjacency matrix A_{ij} computed up to the n -power term of the $e^{A_{ij}}$ series expansion:

$$C_{ij} = e^{A_{ij}} - I \approx C_{ij}(n) = A_{ij} + \frac{(A^2)_{ij}}{2!} + \frac{(A^3)_{ij}}{3!} + \dots + \frac{(A^n)_{ij}}{n!}, \quad (5)$$

where I is the identity matrix. In simple terms, $C_{ij}(n)$ represents a $n!$ -weighted sum of all walks (pathways) of the lengths 1, 2, 3... n joining a pair of network nodes with indices (i, j) . In this study, the matrix exponential is calculated for $n \leq 7$ using sparse matrix multiplication algorithms as available with the CSPARSE package [42]. Thereby, computational costs for iterative compilation of a $C_{ij}(n = 7)$ matrix with 7018 diagonal elements on a Intel Core i5-4590 powered PC amount roughly one hour.

Unweighted adjacency matrices indicate existence $A_{ij} = 1$ of connections between each two nodes (i, j) , but they do not consider intensity of their interactions. In order to account for

biologically relevant differences in strength of network interconnections, weighted adjacency matrices are required. Here, we make strength of network edges dependent on differential gene expression. The basic idea consists of constructing a matrix of differential gene expression weighted interactions which entries are positive numbers $W_{ij} \geq 0$ that have the following interpretation

$$\begin{aligned} W_{ij} < 1 & \text{ interaction between } (i, j) \text{ is downregulated,} \\ W_{ij} = 1 & \text{ interaction between } (i, j) \text{ is unchanged,} \\ W_{ij} > 1 & \text{ interaction between } (i, j) \text{ is upregulated.} \end{aligned} \tag{6}$$

Since interactions between neighbor nodes are defined on network edges, we want to define a mapping function which maps differential gene expression of each two neighbor nodes on their interlinking edge. Furthermore, this mapping should consider that communicability of a single network link is critically dependent on intensities of interlinked nodes. The above requirements are met by the following mapping function which is further termed as the minimum metric (shortly, min-metric):

$$W_{ij} = 2^{\min(\lambda_i, \lambda_j)}. \tag{7}$$

In addition, we introduce an alternative average metric (shortly, avg-metric)

$$W_{ij} = 2^{0.5(\lambda_i + \lambda_j)} \tag{8}$$

that allows to account for a global trend of gene regulation in the entire pathway. In analogy to Eq (5), the matrix of differential gene expression weighted communicability G_{ij} is defined as

$$G_{ij} = e^{W_{ij}} - I \approx G_{ij}(n) = W_{ij} + \frac{(W^2)_{ij}}{2!} + \frac{(W^3)_{ij}}{3!} + \dots + \frac{(W^n)_{ij}}{n!}. \tag{9}$$

Finally, in order to avoid artificial overweighting of well-studied genes with high number of interacting neighbors, we introduce the normalized total communicability (NTC) matrix $D_{ij}(n)$ where non-zero entries are defined by the ratio of differential gene expression weighted to unweighted communicability:

$$D_{ij}(n) = \frac{G_{ij}(n)}{C_{ij}(n)}. \tag{10}$$

Consequently, entries of $D_{ij}(n)$ matrices indicate relative changes in total information flow between each the two network nodes (i, j) joined by pathways of the maximum length n independently on the total number of these pathways.

Results

Starting from 74538 directed interactions between 7018 network nodes, NTC matrices of multistep pathways are computed iteratively as described above (Eqs (5)–(10)). Complete lists of weighted and unweighted pairwise interactions (i.e., 1st order adjacency matrices) for tumor/norm, norm/tumor and ‘random expression’ samples are in S3 Table. With increasing pathway lengths, communicability matrices become densely populated. As shown in Fig 3, the occupancy of communicability matrices (i.e., the ratio of non-zero matrix entries to the dimension of the fully occupied matrix 7018^2) displays a particularly rapid increase from 0.15% to 53% at $n = 4$ and saturates around 70%. This means that the majority of network nodes are interconnected via $n \geq 4$ distant pathways.

n	# interactions	%
1	74538	0.15
2	1525421	3.10
3	11664497	23.68
4	26066024	52.92
5	31761857	64.49
6	32969419	66.94
7	33170092	67.35

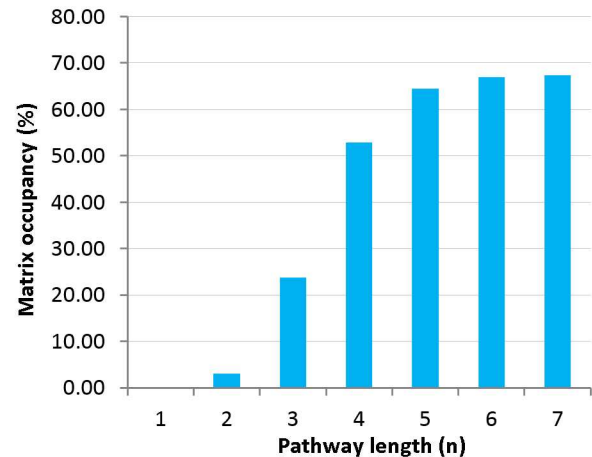


Fig 3. Total number of non-zero entries and percentage of occupancy of communicability matrices as a function of the maximum pathway length $n = 1 - 7$.

doi:10.1371/journal.pone.0170953.g003

Differences between cancer and non-cancer interactomes are first studied using the more restrictive min-metric (Eq (7)). For this purpose, seven NTC matrices $D_{ij}(n = 1 - 7)$ are computed for BRCA/COAD/GBM cancer interactomes. To analyze dependency of NTC on gene scoring (i.e., gene expression), complementary ‘non-cancer’ NTC matrices are assembled by resorting the gene lists in reverse or random orders, i.e.,

$$\begin{aligned} \text{Reverse : } & \lambda_i \leftrightarrow \lambda_{N-i}, \\ \text{Random : } & \lambda_\alpha \leftrightarrow \lambda_\beta, \end{aligned} \tag{11}$$

where $1 \leq \alpha \neq \beta \leq N$ are two unequal random indices of differentially expressed genes in sorted BRCA/COAD/GBM gene lists.

To assess global differences between cancer and non-cancer interactomes, the average of all $D_{ij}(n)$ entries

$$\bar{D}(n) = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M D_{ij}(n), \tag{12}$$

as well as the fraction of enhanced communicability $r(n)$ as a function of the maximum pathway length ($n = 1 - 7$) are computed:

$$r(n) = \left(\sum_{k=1}^K \sum_{l=1}^L D_{kl}(n) \right) \left(\sum_{i=1}^N \sum_{j=1}^M D_{ij}(n) \right)^{-1}, \tag{13}$$

where (k, l) and (i, j) denote indices of elevated (i.e., $D_{kl}(n) > 1$) and all NTC entries, respectively. Fig 4 shows plots of $\bar{D}(n)$ and $r(n)$ for cancer and non-cancer networks. Remarkably, the average NTC in the range of $n \geq 4$ distant pathways exhibits a persistent increase only in cancer interactomes. In contrast, average NTC of all non-cancer networks declines with an increasing pathway length. Difference between cancer and non-cancer NTC is also visible in the fraction of elevated NTC as a function of the maximum pathway length $r(n)$, which shows

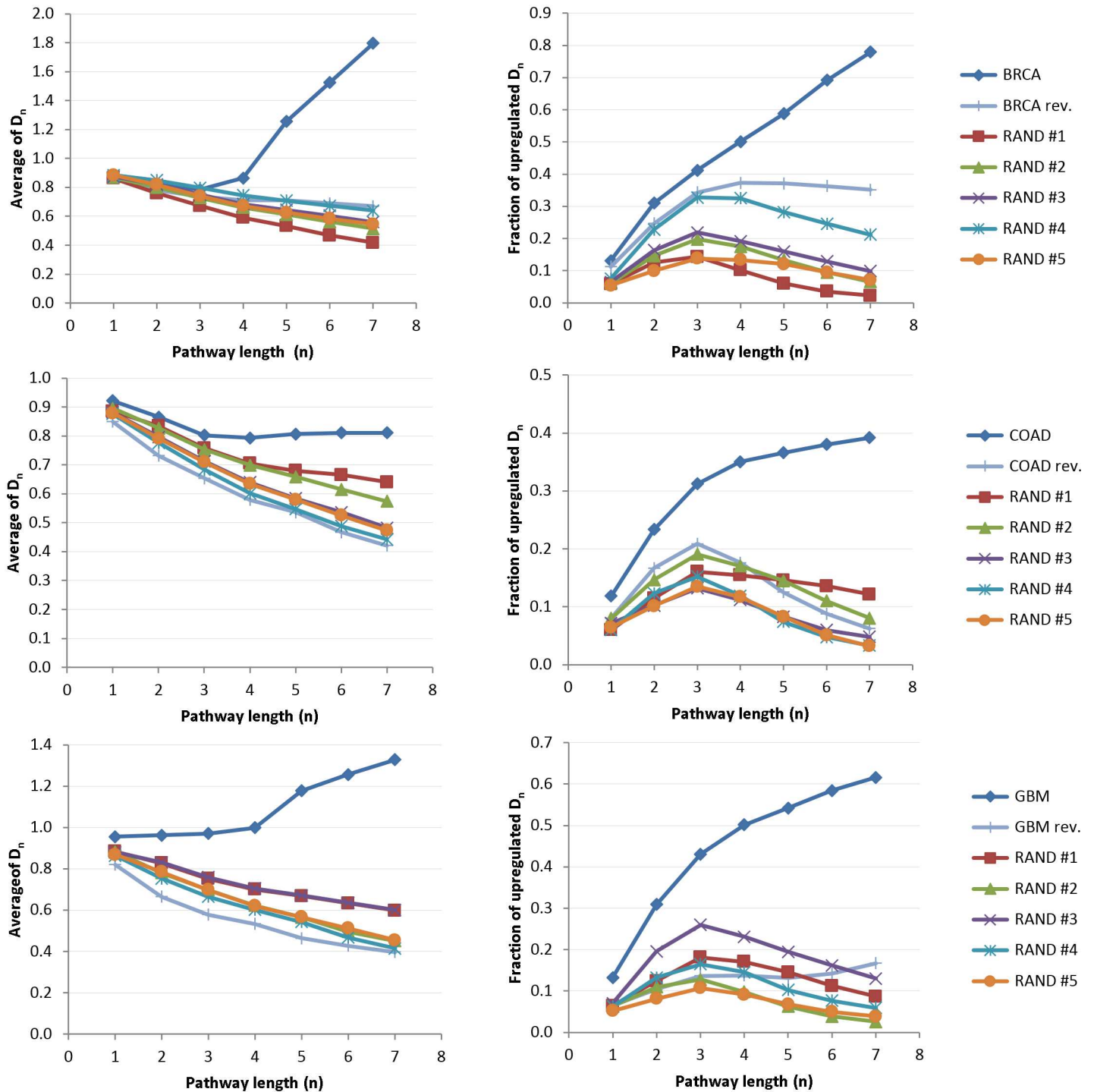


Fig 4. Statistics of total communicability matrices computed using the min-metric. Left column: average NTC as a function of the maximum pathway length ($n = 1 - 7$). Right column: fraction of enhanced network communicability (Eq (13)) as a function of the maximum pathway length ($n = 1 - 7$). In contrast to normal and randomly weighted networks, BRCA/COAD/GBM cancer networks exhibit elevated long-range NTC.

doi:10.1371/journal.pone.0170953.g004

more rapid growth in cancer than in the reference non-cancer networks. Similar patterns of elevated long-range NTC in cancer interactomes are also observed when using the avg-metric (Eq (8)). To compare NTC of cancer and non-cancer networks simultaneously in both metrics, diagonal (θ) and cumulative off-diagonal (ξ) communicabilities in min- and avg-metric are computed as follows

$$\begin{aligned} \theta_i^{\min}(n) &= \log(D_{ii}^{\min}(n)) & \zeta_i^{\min}(n) &= \log\left(\sum_{j \neq i}^M D_{ij}^{\min}(n)\right), \\ \theta_i^{\text{avg}}(n) &= \log(D_{ii}^{\text{avg}}(n)) & \zeta_i^{\text{avg}}(n) &= \log\left(\sum_{j \neq i}^M D_{ij}^{\text{avg}}(n)\right). \end{aligned} \tag{14}$$

Fig 5 shows diagonal and off-diagonal $n \leq 7$ gene communicabilities of BRCA/COAD/GBM vs randomly weighted interactomes as two-dimensional distributions. Significance of the differences between cancer and non-cancer communicabilities in $(\theta^{\min}, \theta^{\text{avg}})$ and $(\zeta^{\min}, \zeta^{\text{avg}})$ representations is confirmed by the two-dimensional Kolmogorov-Smirnov test with significance level $p < 0.001$.

To determine whether elevated communicability of different cancer interactomes arises from enhancement of common genes, the impact of simulated gene inhibition on the above statistical features of NTC matrices (Eqs (12)–(13)) is simulated. For this purpose, a cut set of common BRCA/COAD/GBM genes with high NTC in both min- and avg-metric is computed using an iterative procedure, as shown in Fig 6. Starting with the initial set of 530 common BRCA/COAD/GBM genes, the smallest subset of 90 genes is identified whose simulated inhibition is sufficient to decrease the difference between NTC of cancer and non-cancer interactomes, see Fig 5 (orange labels). Subsequent visualization and ontological analysis using STRING reveal association of these tightly interlinked genes with enhanced cell division, DNA replication, cellular stress response and other cancer related functional categories, see Fig 7 and S4 Table. In addition to common genes, there are cancer-type specific genes with high NTC that appear to group in separate clusters in min-avg diagrams. 76 COAD specific genes, indicated in Fig 5 with green labels, build a prominent cluster with particularly high NTC values in avg-metric. These 76 genes are enriched in *Hedgehog*, *Hippo* and *Wnt* pathways which are known to promote Epithelial-to-Mesenchymal Transition (EMT) and metastatic cell transformation [43], see Fig 8 and S5 Table.

Since our above simulations indicate a high level of robustness of cancer networks with respect to inhibition of a relatively small number of genes, we are interested in assessing the inhibitory effects of other prominent gene signatures. For this purpose, we examine a list of 33 genes with a high differential entropy in bladder cancer highlighted in [31]. Remarkably, 12 of 33 bladder cancer genes from [31] are also present in our list of 90 common BRCA/COAD/GBM genes. According to the hypergeometric test $HGT(7018, 90, 33, 12) = 2.6e-15$, this overlap is statistically significant, see S6 Table. However, simulated inhibition of these 33 genes turned out to not be sufficient for the suppression of elevated NTC. Fig 9 shows the results of simulated inhibition 33 bladder cancer genes from [31] and 90 common BRCA/COAD/GBM genes identified in this work.

Discussion

Network-based approaches to mining omics data are increasingly popular. However, consistent modeling of biological networks remains a challenging task and requires the consideration of numerous factors whose impact on simulation results is still controversially debated in the

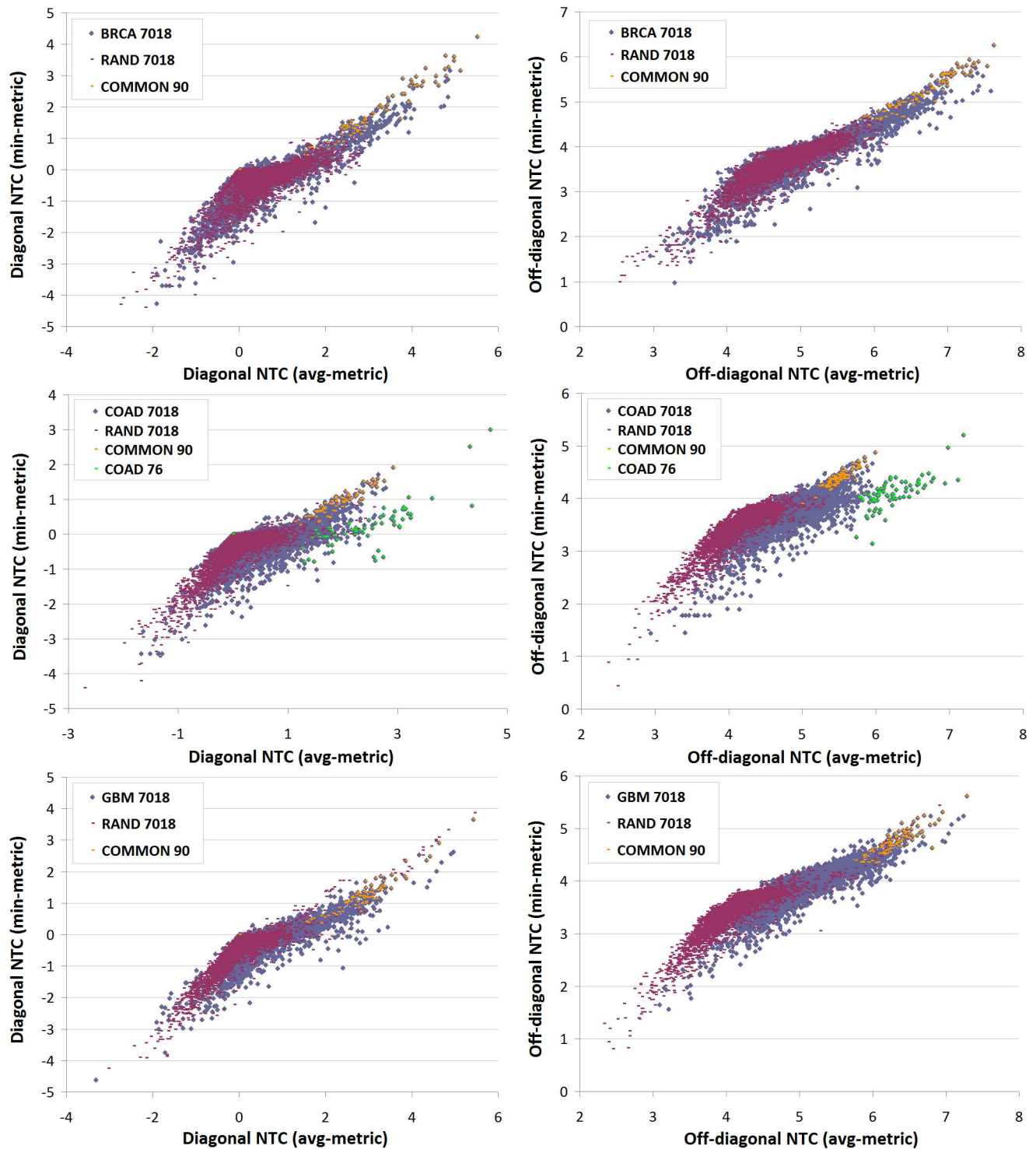


Fig 5. Diagonal and off-diagonal communicability of tumor (cyan) vs random (bordeaux) interactomes in min- and avg-metric. Each point represents log-sum of total diagonal (left column) re. off-diagonal (right column) communicability of the i -th network node to itself and remote neighbors via all walks of the length ($n \leq 7$), respectively. Orange labels indicate a subset of 90 common genes that are associated with elevated communicability in BRCA/COAD/GBM interactomes. Green labels in the COAD plot indicate a fraction of 76 COAD specific EMT-related genes with particularly high communicability in avg-metric.

doi:10.1371/journal.pone.0170953.g005

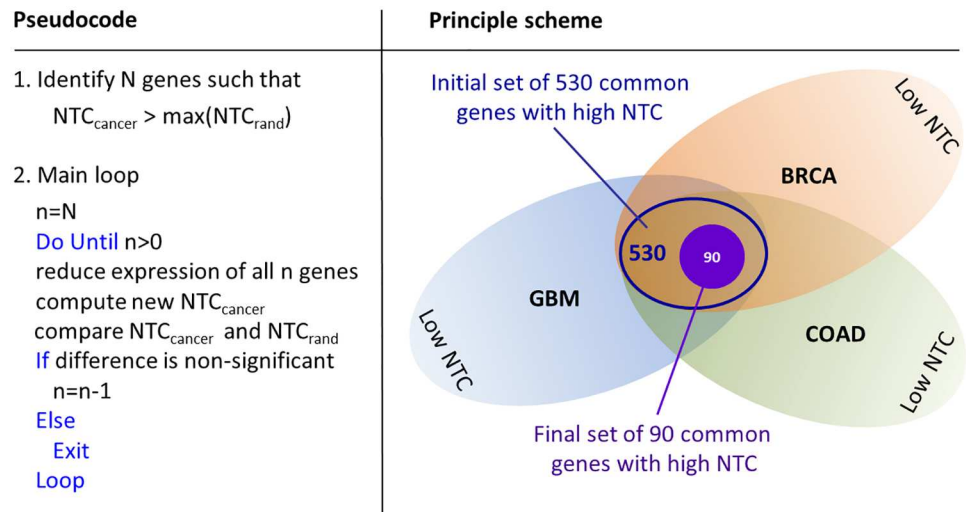


Fig 6. Identification of the cut set of genes associated with elevated NTC in BRCA/COAD/GBM cancer samples. First, an initial cut set of 530 common BRCA/COAD/GBM genes with elevated NTC is estimated. By iterative reduction of the initial gene set, 90 genes are identified whose simulated inhibition is sufficient to level down the difference between NTC of cancer and non-cancer interactomes.

doi:10.1371/journal.pone.0170953.g006

literature. These factors include the role of a particular network topology, directionality and size as well as the choice of appropriate gene proximity metrics and numerical scores. In this work, we focused on the construction and evaluation of novel descriptors for measurement of network information flow. We let other issues remain widely unaddressed, assuming that simulation results obtained with different network topologies should be, in general, convergent.

To account for the potential bias of degree-based network features, we introduced a degree-independent measure of information flow—the normalized total communicability (NTC). NTC relies on a well-known concept of network topology characterization by means of the n -power adjacency matrices, whose entries indicate the total number of unweighted walks of the maximum length n between each two network nodes. In our approach, adjacency matrices of unweighted network topology are used for normalization of differential gene expression weighted walks. Consequently, NTC does not explicitly depend on node degrees, but rather serves as an integrative measure of up- or downregulation of all pathways of the maximum length n joining each two network nodes.

Similar to other works, we use public databases on pairwise protein-protein and gene regulatory interactions to compile the basic network topology. However, here we rely on a subset of directed interactions that naturally restrict the emergence of loops to $n \geq 3$ network steps. Our simulation results show that elevated long-range communicability of different cancer networks is largely caused by circulation of information flow within compact subnetworks of tightly interlinked genes. In networks with non-directed interactions, this feature might be missing.

Our simulations indicate a high level of robustness of cancer networks with respect to inhibition of a low number of genes-proteins. Simulated inhibition of a few dozen genes, including a hit list of 33 bladder cancer genes from [31], was not sufficient to suppress elevation of long-

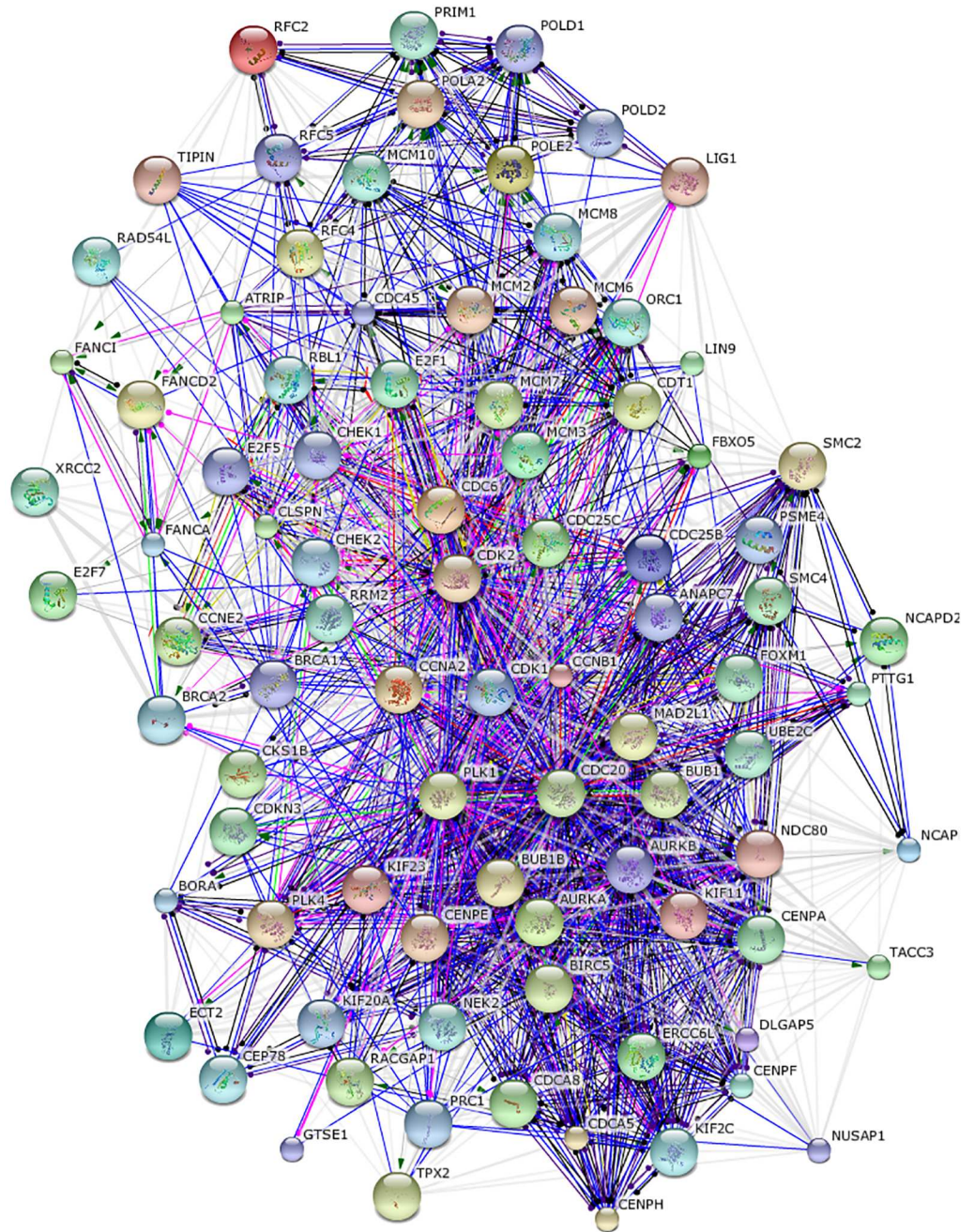


Fig 7. Subnetwork of 90 common BRCA/COAD/GBM genes with high NTC, cf. Fig 5 (orange labels).

doi:10.1371/journal.pone.0170953.g007

range NTC. Despite the fact that elevated NTC arises in different cancer interactomes from heterogeneous gene expression profiles, we identified a subset of 90 common BRCA/COAD/GBM genes whose simulated inhibition is capable of reducing differences between NTC of cancer and non-cancer interactomes. These genes turn out to be associated with cancer-related ontological categories, including enhanced cell division, DNA replication, elevated energy demand and cellular stress response. We conclude that enhanced NTC reflects correlated

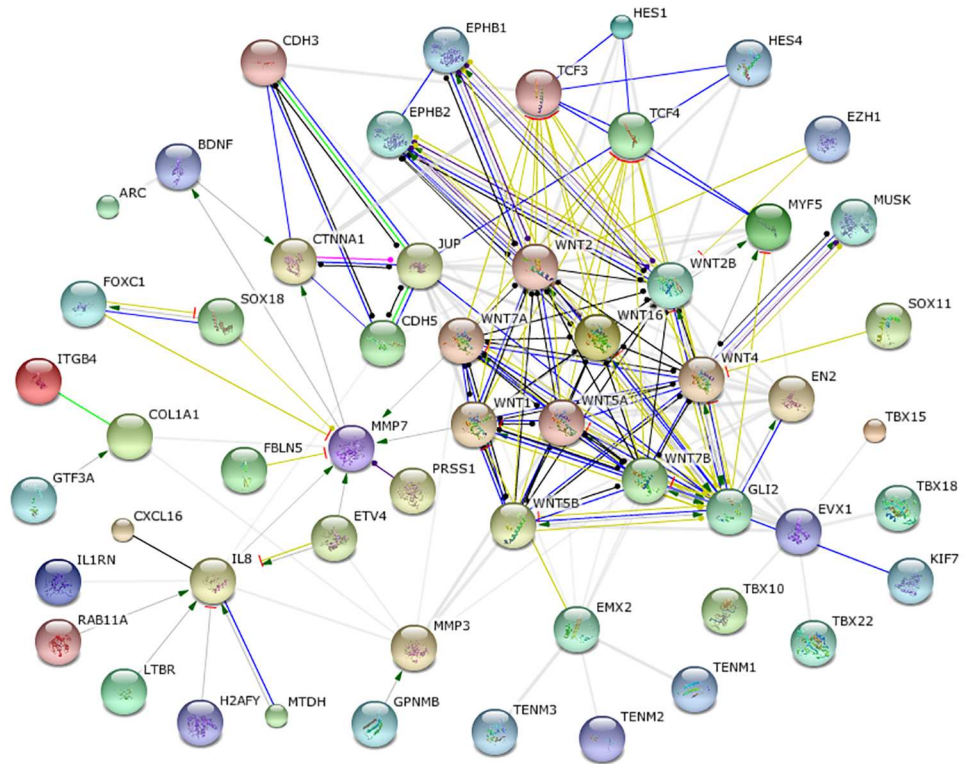


Fig 8. Subnetwork of 76 COAD specific genes with high NTC in the avg-metric, cf. Fig 5 (green labels).

doi:10.1371/journal.pone.0170953.g008

activity of genes whose coordination is required for maintenance of sustained proliferation and replication of all tumor cells. In other words, elevated long-range NTC represents a graph-theoretical hallmark of cancer networks. Under the assumption of gradual elevation of NTC in course of cancer development, an abnormal increase of NTC can serve as an early marker of malignant cell transformation. Further investigations of normally proliferating and cancer cells at different stages of disease development are required to prove this assumption and to define reliable diagnostic measures.

While focusing on construction of a feasible graph-theoretical formalism for gene expression weighted network modeling, this work does not go into the discussion of biological mechanisms of cancer network rewiring and regulation. Different biological processes on single gene, chromosome and whole genome level including gene mutations, changes in gene copy number, chromotripsis are known to accompany malignant cell transformation [44]. Consideration of this layer of information will be an important subject for future research.

Finally, our graph-theoretical framework is of potential interest for a broad spectrum of network perturbation problems such as single or multiple gene-protein activation, inhibition or malfunction due to the impact of mutations or interactions with pharmaceutical drugs.

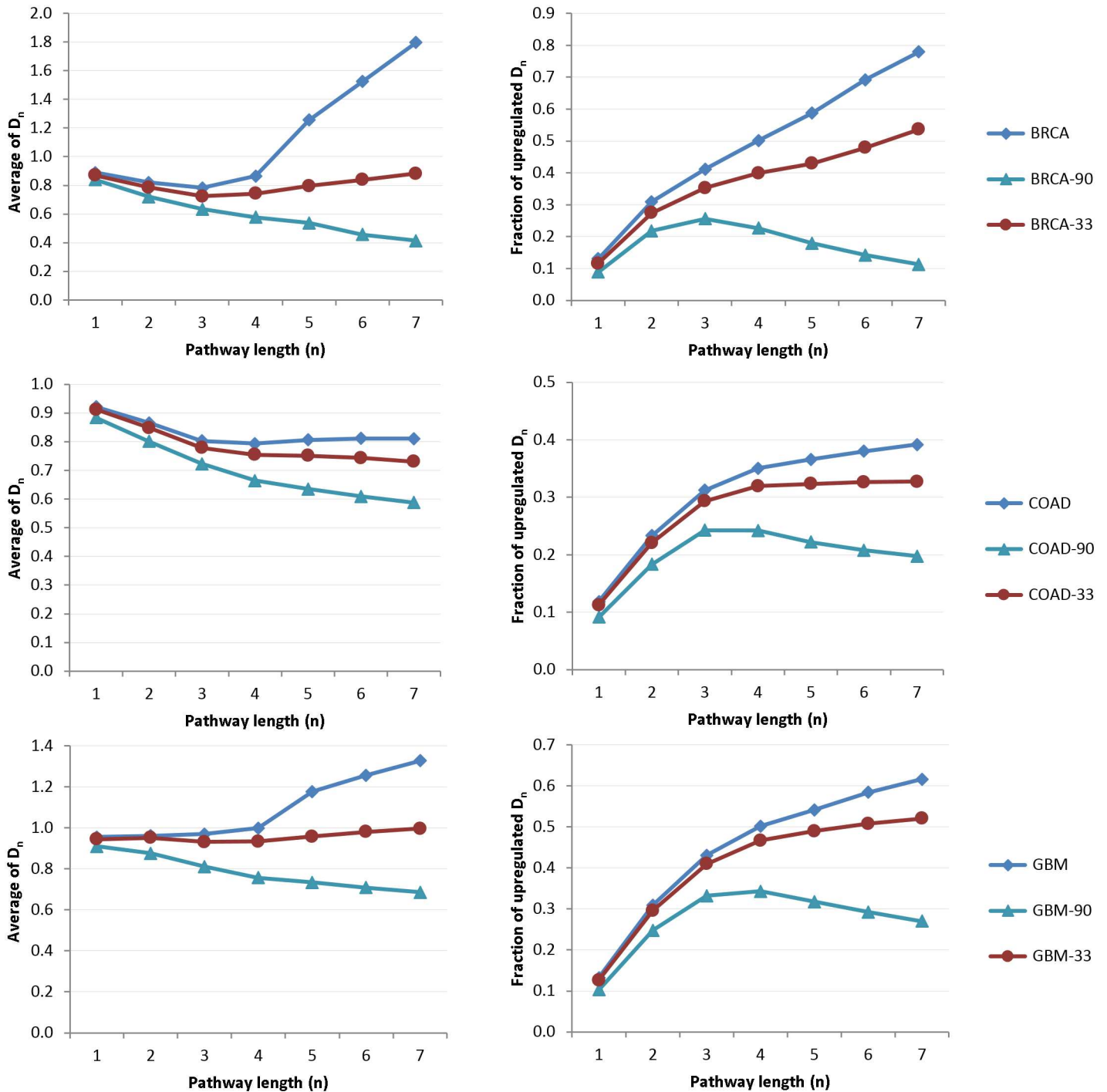


Fig 9. Effects of simulated inhibition of 90 common BRCA/COAD/GBM genes with high NTC vs 33 bladder cancer genes with high differential entropy from [31], see full gene lists in S4 Table. Inhibition of 33 high-score targets from [31] moderately decreases global communicability features of cancer networks. However, it is obviously not sufficient to suppress their elevation with increasing pathway length. In contrast, simulated inhibition of our 90 targets results in decreasing the difference between NTC features of cancer and non-cancer interactomes.

doi:10.1371/journal.pone.0170953.g009

Supporting Information

S1 Table. Lists of TCGA BRCA, COAD, GBM tumor/norm samples used in this study.
(XLSX)

S2 Table. Sorted lists of average differential gene expression (ADGE) values of TCGA BRCA/COAD/GBM tumor/norm (TN), norm/tumor (NT) (i.e., reversed TN) and 'random expression' data.
(XLSX)

S3 Table. Lists of 74538 unweighted and gene expression weighted pairwise interactions (i.e., 1st order adjacency matrices) computed on the basis of TCGA BRCA/COAD/GBM TN, NT and 'random expression' data.
(XLSX)

S4 Table. 90 common BRCA/COAD/GBM genes associated with elevated NTC and their GO enrichment terms.
(XLSX)

S5 Table. 76 COAD-specific genes associated with elevated NTC in the avg-metric and their GO enrichment terms.
(XLSX)

S6 Table. Overlap between 90 common BRCA/COAD/GBM genes with high NTC (see [S4 Table](#)) and 33 genes with the high differential entropy in bladder cancer from [31](Suppl. Table.2).
(XLSX)

Acknowledgments

The author is grateful to Amanda Chase for critically reading the manuscript.

Author Contributions

Conceptualization: E.G.

Data curation: E.G.

Formal analysis: E.G.

Investigation: E.G.

Methodology: E.G.

Project administration: E.G.

Software: E.G.

Validation: E.G.

Visualization: E.G.

Writing – original draft: E.G.

Writing – review & editing: E.G.

References

1. Marusyk A, Polyak K. Tumor heterogeneity: Causes and consequences. *Biochim Bioph Acta—Reviews on Cancer*. 2010; 1805(1):105–117. doi: [10.1016/j.bbcan.2009.11.002](https://doi.org/10.1016/j.bbcan.2009.11.002) PMID: [19931353](https://pubmed.ncbi.nlm.nih.gov/19931353/)
2. Hanahan D, Weinberg R. Hallmarks of cancer: the next generation. *Cell*. 2011; 144(5):646–674. doi: [10.1016/j.cell.2011.02.013](https://doi.org/10.1016/j.cell.2011.02.013) PMID: [21376230](https://pubmed.ncbi.nlm.nih.gov/21376230/)
3. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási AL. The large-scale organization of metabolic networks. *Nature*. 2000; 407(6804):651–654. doi: [10.1038/35036627](https://doi.org/10.1038/35036627) PMID: [11034217](https://pubmed.ncbi.nlm.nih.gov/11034217/)
4. Barabási AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nature reviews Genetics*. 2004; 5(2):101–113. doi: [10.1038/nrg1272](https://doi.org/10.1038/nrg1272) PMID: [14735121](https://pubmed.ncbi.nlm.nih.gov/14735121/)
5. Christensen C, Albert R. Using graph concepts to understand the organization of complex systems. *Int J Bifurcation and Chaos*. 2006; 17:2201. doi: [10.1142/S021812740701835X](https://doi.org/10.1142/S021812740701835X)
6. Hanisch D, Zien A, Zimmer R, Lengauer T. Co-clustering of biological networks and gene expression data. *Bioinformatics*. 2002; 1(18):S145–54. doi: [10.1093/bioinformatics/18.suppl_1.S145](https://doi.org/10.1093/bioinformatics/18.suppl_1.S145) PMID: [12169542](https://pubmed.ncbi.nlm.nih.gov/12169542/)
7. Zhang B, Horvath S. A General Framework for Weighted Gene Co-Expression Network Analysis. *Stat Appl Genet Mol Biol*. 2005; 4(1):17. doi: [10.2202/1544-6115.1128](https://doi.org/10.2202/1544-6115.1128) PMID: [16646834](https://pubmed.ncbi.nlm.nih.gov/16646834/)
8. Horvath S, Dong J. Geometric interpretation of gene coexpression network analysis. *PLoS Computational Biology*. 2008; 4(8):24–26. doi: [10.1371/journal.pcbi.1000117](https://doi.org/10.1371/journal.pcbi.1000117) PMID: [18704157](https://pubmed.ncbi.nlm.nih.gov/18704157/)
9. Bader GG, Hogue CC. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*. 2003; 4:2. doi: [10.1186/1471-2105-4-2](https://doi.org/10.1186/1471-2105-4-2) PMID: [12525261](https://pubmed.ncbi.nlm.nih.gov/12525261/)
10. Tornow S, Mewes HW. Functional modules by relating protein interaction networks and gene expression. *Nucleic Acids Research*. 2003; 31(21):6283–6289. doi: [10.1093/nar/gkg838](https://doi.org/10.1093/nar/gkg838) PMID: [14576317](https://pubmed.ncbi.nlm.nih.gov/14576317/)
11. Tanay A, Sharan R, Kupiec M, Shamir R. Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous genomewide data. *Proceedings of the National Academy of Sciences of the United States of America*. 2004; 101(9):2981–2986. doi: [10.1073/pnas.0308661100](https://doi.org/10.1073/pnas.0308661100) PMID: [14973197](https://pubmed.ncbi.nlm.nih.gov/14973197/)
12. Chua HN, Sung WK, Wong L. Exploiting indirect neighbours and topological weight to predict protein function from protein-protein interactions. *Bioinformatics (Oxford, England)*. 2006; 22(13):1623–30. doi: [10.1093/bioinformatics/btl145](https://doi.org/10.1093/bioinformatics/btl145) PMID: [16632496](https://pubmed.ncbi.nlm.nih.gov/16632496/)
13. Ulitsky I, Shamir R. Identification of functional modules using network topology and high-throughput data. *BMC Systems Biology*. 2007; 1:8. doi: [10.1186/1752-0509-1-8](https://doi.org/10.1186/1752-0509-1-8) PMID: [17408515](https://pubmed.ncbi.nlm.nih.gov/17408515/)
14. Ideker T, Sharan R. Protein networks in disease. *Genome Res*. 2008; 18:644–652. doi: [10.1101/gr.071852.107](https://doi.org/10.1101/gr.071852.107) PMID: [18381899](https://pubmed.ncbi.nlm.nih.gov/18381899/)
15. Liu G, Wong L, Chua HN. Complex discovery from weighted PPI networks. *Bioinformatics*. 2009; 25(15):1891–1897. doi: [10.1093/bioinformatics/btp311](https://doi.org/10.1093/bioinformatics/btp311) PMID: [19435747](https://pubmed.ncbi.nlm.nih.gov/19435747/)
16. Li X, Wu M, Kwok CK, Ng SK. Computational approaches for detecting protein complexes from protein interaction networks: a survey. *BMC Genomics*. 2010; 11 Suppl 1:S3. doi: [10.1186/1471-2164-11-S1-S3](https://doi.org/10.1186/1471-2164-11-S1-S3) PMID: [20158874](https://pubmed.ncbi.nlm.nih.gov/20158874/)
17. Liu H, Su J, Li J, Liu H, Lv J, Li B, et al. Prioritizing cancer-related genes with aberrant methylation based on a weighted protein-protein interaction network. *BMC Systems Biology*. 2011; 5(1):158. doi: [10.1186/1752-0509-5-158](https://doi.org/10.1186/1752-0509-5-158) PMID: [21985575](https://pubmed.ncbi.nlm.nih.gov/21985575/)
18. Feng J, Jiang R, Jiang T. A max-flow-based approach to the identification of protein complexes using protein interaction and microarray data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. 2011; 8(3):621–634. doi: [10.1109/TCBB.2010.78](https://doi.org/10.1109/TCBB.2010.78) PMID: [20733237](https://pubmed.ncbi.nlm.nih.gov/20733237/)
19. Milenković T, Memišević V, Bonato A, Pržulj N. Dominating biological networks. *PLoS ONE*. 2011; 6(8):28–32. doi: [10.1371/journal.pone.0023016](https://doi.org/10.1371/journal.pone.0023016) PMID: [21887225](https://pubmed.ncbi.nlm.nih.gov/21887225/)
20. Cho DY, Kim YA, Przytycka TM. Chapter 5: Network Biology Approach to Complex Diseases. *PLoS Computational Biology*. 2012; 8(12):e1002820. doi: [10.1371/journal.pcbi.1002820](https://doi.org/10.1371/journal.pcbi.1002820) PMID: [23300411](https://pubmed.ncbi.nlm.nih.gov/23300411/)
21. Guney E, Oliva B. Exploiting Protein-Protein Interaction Networks for Genome-Wide Disease-Gene Prioritization. *PLoS ONE*. 2012; 7(9):e43557. doi: [10.1371/journal.pone.0043557](https://doi.org/10.1371/journal.pone.0043557) PMID: [23028459](https://pubmed.ncbi.nlm.nih.gov/23028459/)
22. Yu D, Kim M, Xiao G, Hwang TH. Review of biological network data and its applications. *Genomics & Informatics*. 2013; 11(4):200–210. doi: [10.5808/GI.2013.11.4.200](https://doi.org/10.5808/GI.2013.11.4.200)
23. Pandey G, Arora S, Manocha S, Whalen S. Enhancing the functional content of eukaryotic protein interaction networks. *PLoS ONE*. 2014; 9(10):e109130. doi: [10.1371/journal.pone.0109130](https://doi.org/10.1371/journal.pone.0109130) PMID: [25275489](https://pubmed.ncbi.nlm.nih.gov/25275489/)

24. Tuck DP, Kluger HM, Kluger Y. Characterizing disease states from topological properties of transcriptional regulatory networks. *BMC Bioinformatics*. 2006; 7:236. doi: [10.1186/1471-2105-7-236](https://doi.org/10.1186/1471-2105-7-236) PMID: [16670008](https://pubmed.ncbi.nlm.nih.gov/16670008/)
25. Platzer A, Perco P, Lukas A, Mayer B. Characterization of protein-interaction networks in tumors. *BMC Bioinformatics*. 2007; 8:224. doi: [10.1186/1471-2105-8-224](https://doi.org/10.1186/1471-2105-8-224) PMID: [17597514](https://pubmed.ncbi.nlm.nih.gov/17597514/)
26. Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, et al. Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nature biotechnology*. 2009; 27(2):199–204. doi: [10.1038/nbt.1522](https://doi.org/10.1038/nbt.1522) PMID: [19182785](https://pubmed.ncbi.nlm.nih.gov/19182785/)
27. Komurov K, Ram PT. Patterns of human gene expression variance show strong associations with signaling network hierarchy. *BMC Systems Biology*. 2010; 4(1):154. doi: [10.1186/1752-0509-4-154](https://doi.org/10.1186/1752-0509-4-154) PMID: [21073694](https://pubmed.ncbi.nlm.nih.gov/21073694/)
28. Weiss JN, Karma A, MacLellan WR, Deng M, Rau CD, Rees CM, et al. “Good enough solutions” and the genetics of complex diseases. *Circulation research*. 2012; 111(4):493–504. doi: [10.1161/CIRCRESAHA.112.269084](https://doi.org/10.1161/CIRCRESAHA.112.269084) PMID: [22859671](https://pubmed.ncbi.nlm.nih.gov/22859671/)
29. Schramm G, Kannabiran N, König R. Regulation patterns in signaling networks of cancer. *BMC Systems Biology*. 2010; 4(1):162. doi: [10.1186/1752-0509-4-162](https://doi.org/10.1186/1752-0509-4-162) PMID: [21110851](https://pubmed.ncbi.nlm.nih.gov/21110851/)
30. Albert R, DasGupta B, Mobasher N. Topological implications of negative curvature for biological and social networks. *Phys Rev E Stat Nonline Soft Matter Phys*. 2014; 89:032811. doi: [10.1103/PhysRevE.89.032811](https://doi.org/10.1103/PhysRevE.89.032811) PMID: [24730903](https://pubmed.ncbi.nlm.nih.gov/24730903/)
31. West J, Bianconi G, Severini S, Teschendorff AE. Differential network entropy reveals cancer system hallmarks. *Scientific reports*. 2012; 2:802. doi: [10.1038/srep00802](https://doi.org/10.1038/srep00802) PMID: [23150773](https://pubmed.ncbi.nlm.nih.gov/23150773/)
32. Banerji CRS, Miranda-Saavedra D, Severini S, Widschwendter M, Enver T, Zhou JX, et al. Cellular network entropy as the energy potential in Waddington’s differentiation landscape. *Scientific Reports*. 2013; 3:25–27. doi: [10.1038/srep03039](https://doi.org/10.1038/srep03039) PMID: [24154593](https://pubmed.ncbi.nlm.nih.gov/24154593/)
33. Jonsson PF, Bates PA. Global topological features of cancer proteins in the human interactome. *Bioinformatics (Oxford, England)*. 2006; 22(18):2291–2297. doi: [10.1093/bioinformatics/btl390](https://doi.org/10.1093/bioinformatics/btl390) PMID: [16844706](https://pubmed.ncbi.nlm.nih.gov/16844706/)
34. Milenkovic T, Memisevic V, Ganesan AK, Przulj N. Systems-Level Cancer Gene Identification from Protein Interaction Network Topology Applied to Melanogenesis-Related Functional Genomics Data. *Journal of The Royal Society Interface*. 2010; 7(44):423–437. doi: [10.1098/rsif.2009.0192](https://doi.org/10.1098/rsif.2009.0192) PMID: [19625303](https://pubmed.ncbi.nlm.nih.gov/19625303/)
35. Islam MF, Hoque MM, Banik RS, Roy S, Sumi SS, Hassan FMN, et al. Comparative analysis of differential network modularity in tissue specific normal and cancer protein interaction networks. *Journal of clinical bioinformatics*. 2013; 3(1):19. doi: [10.1186/2043-9113-3-19](https://doi.org/10.1186/2043-9113-3-19) PMID: [24093757](https://pubmed.ncbi.nlm.nih.gov/24093757/)
36. Rai A, Menon AV, Jalan S. Randomness and preserved patterns in cancer network. *Scientific reports*. 2014; 4:6368. doi: [10.1038/srep06368](https://doi.org/10.1038/srep06368) PMID: [25220184](https://pubmed.ncbi.nlm.nih.gov/25220184/)
37. Guney E, Oliva B. Analysis of the robustness of network-based disease-gene prioritization methods reveals redundancy in the human interactome and functional diversity of disease-genes. *PLoS ONE*. 2014; 9(4):e94686. doi: [10.1371/journal.pone.0094686](https://doi.org/10.1371/journal.pone.0094686) PMID: [24733074](https://pubmed.ncbi.nlm.nih.gov/24733074/)
38. Hakes L, Pinney JW, Robertson DL, Lovell SC. Protein-protein interaction networks and biology—what’s the connection? *Nature biotechnology*. 2008; 26(1):69–72. doi: [10.1038/nbt0108-69](https://doi.org/10.1038/nbt0108-69) PMID: [18183023](https://pubmed.ncbi.nlm.nih.gov/18183023/)
39. Schaefer MH, Serrano L, Andrade-Navarro Ma. Correcting for the study bias associated with protein-protein interaction measurements reveals differences between protein degree distributions from different cancer types. *Frontiers in Genetics*. 2015; 6(August):260. doi: [10.3389/fgene.2015.00260](https://doi.org/10.3389/fgene.2015.00260) PMID: [26300911](https://pubmed.ncbi.nlm.nih.gov/26300911/)
40. Estrada E, Rodríguez-Velázquez JA. Subgraph centrality in complex networks. *Physical Review E*. 2005; 71(5):056103. doi: [10.1103/PhysRevE.71.056103](https://doi.org/10.1103/PhysRevE.71.056103)
41. Benzi M, Klymko C. Total communicability as a centrality measure. *Journal of Complex Networks*. 2013; 1(2):124–149. doi: [10.1093/comnet/cnt007](https://doi.org/10.1093/comnet/cnt007)
42. Davis TA. *Direct Methods for Sparse Linear Systems (Fundamentals of Algorithms 2)*. Philadelphia, PA, USA: SIAM; 2006.
43. Lamouille S, Xu J, Derynck R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol*. 2014; 15(3):178–196. doi: [10.1038/nrm3758](https://doi.org/10.1038/nrm3758) PMID: [24556840](https://pubmed.ncbi.nlm.nih.gov/24556840/)
44. Heng H, Liu G, Stevens J, Bremer S, Ye K, Abdallah B, et al. Decoding the genome beyond sequencing: the new phase of genomic research. *Genomics*. 2011; 98(4):242–252. doi: [10.1016/j.ygeno.2011.05.008](https://doi.org/10.1016/j.ygeno.2011.05.008) PMID: [21640814](https://pubmed.ncbi.nlm.nih.gov/21640814/)