THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

Modular Reorganization of Signaling Networks during the Development of Colon Adenoma and Carcinoma

Published as part of The Journal of Physical Chemistry virtual special issue "Ruth Nussinov Festschrift".

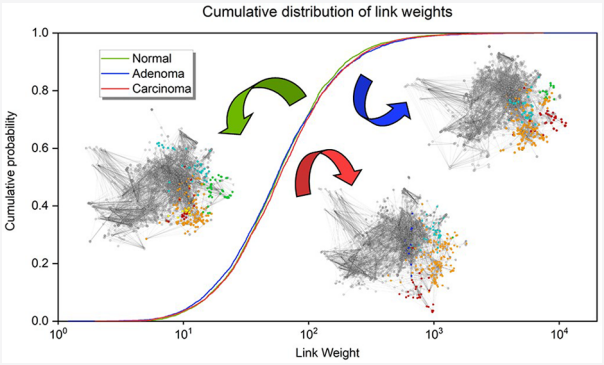
Klára Schulc, Zsolt T. Nagy, Sebestyén Kamp, János Molnár, Daniel V. Veres, Peter Csermely,* and Borbála M. Kovács

Cite This: *J. Phys. Chem. B* 2021, 125, 1716–1726

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Network science is an emerging tool in systems biology and oncology, providing novel, system-level insight into the development of cancer. The aim of this project was to study the signaling networks in the process of oncogenesis to explore the adaptive mechanisms taking part in the cancerous transformation of healthy cells. For this purpose, colon cancer proved to be an excellent candidate as the preliminary phase, and adenoma has a long evolution time. In our work, transcriptomic data have been collected from normal colon, colon adenoma, and colon cancer samples to calculating link (i.e., network edge) weights as approximative proxies for protein abundances, and link weights were included in the Human Cancer Signaling Network. Here we show that the adenoma phase clearly differs from the normal and cancer states in terms of a more scattered link weight distribution and enlarged network diameter. Modular analysis shows the rearrangement of the apoptosis- and the cell-cycle-related modules, whose pathway enrichment analysis supports the relevance of targeted therapy. Our work enriches the system-wide assessment of cancer development, showing specific changes for the adenoma state.

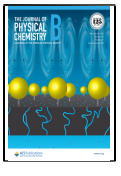


1. INTRODUCTION

Colon cancer is a malignant tumor originating in the large intestine, which is histologically considered to be an adenocarcinoma in 95% of the cases. The relevance of this disease is hard to underestimate, as colorectal cancer (CRC) is the third most common cancer diagnosed in males and the second in females. Although the survival rate is increasing in the United States, the mortality of the disease is still among the leading cancer subtypes according to the GLOBOCAN database of the World Health Organization.[1] Like most civilizational diseases, the etiology of colon cancer is multifactorial, involving both genetic (e.g., familial adenomatous polyposis, Lynch Syndrome) and environmental (lifestyle risk factors such as alcohol and smoking) causes. Benign polyps/adenomas are extremely common, and regretfully, 10% of them transform malignantly. Nearly all CRC develop this way, thus an understanding of the process is essential. There are heterogeneous causes of colon cancer development occurring at the molecular level, involving epigenomic and genomic instabilities, resulting in the deregulations of various signaling pathways involved in cell differentiation and growth.[2] Despite thousands of mutant genes, 15 driver mutations were identified as key features in the pathogenesis of colon cancer.[3] The treatment of the disease mainly includes surgery and chemotherapy,[4] but targeted therapies such as EGFR and VEGFR inhibition and immunotherapy came into view as palliative treatment options.[5]

In recent years, the role of network science in medicine and molecular biology has been growing extensively.[6−9] As technological development allowed the detailed measurement of mRNA expression, networks were constructed to perform data analysis to provide information about signaling in normal and pathological stages of a cell on the system level. Detailed network topology studies have been conducted in which the network diameter was calculated and modular structure was detected via different algorithms.[10−23] In our earlier joint studies with the pioneering expert in the field, Prof. Ruth Nussinov[9,24−26] showed that signaling networks proved to be very valuable tools in characterizing the development of cancer

Received: October 14, 2020
Revised: January 28, 2021
Published: February 9, 2021



https://dx.doi.org/10.1021/acs.jpcb.0c09307
J. Phys. Chem. B 2021, 125, 1716–1726

and in the discovery of new therapeutic modalities. Signaling networks are able to describe the complex crossroads of signal transduction pathways in a clearly understandable way. For this reason, the Human Cancer Signaling Network,[27] which is a signaling network enriched by relating protein−protein interactions, appeared to be an excellent candidate for studying colon cancer, one of the most common oncological diseases in the world.[1]

Network topological analysis focusing on highly overlapping modules has never been applied before to colon cancer data. The modularization approach we used focuses on extensively overlapping and hierarchical modules and provides a powerful way to model cellular processes such as the functional reorganization of protein complexes.[28] This sophisticated separation of modules was required for the results to show the deregulation of the apoptotic process and the increasing activity of the cell cycle.

This article demonstrates that the adenoma network forms a distinguished state between the normal and carcinoma networks. This statement is supported by the findings that the adenoma network has the highest standard deviation in the distribution of the link weights and has the largest diameter. Regarding the changes in the biological functions, the reorganized modules are in the areas of apoptosis regulation and the cell cycle, which are both well-known features of carcinogenesis. In terms of clinical relevance, the Gene Ontology[29,30] pathway enrichment analysis resulted in confirming the strengthening role of the EGFR and VEGFR, whose inhibitors are already in use as palliative treatment options for colorectal cancer. The significant achievement of the work presented here is that it shows that using a highly overlapping modularization method enables us to model the construction of biological systems, opening new possibilities to understanding the pathogenesis of cancer.

2. MATERIALS AND METHODS

2.1. Concept of the Analysis. Building and analyzing networks from the significant fold changes between the expression levels of two biological conditions is a common method. However, with the help of the mRNA abundance values, each three stages of carcinogenesis (normal, adenoma, and carcinoma) can be modeled individually. Moreover, with the use of a standard, previously described network model (i.e., The Human Cancer Signaling Network,[27] see Materials and Methods) a complex picture can be used to represent the biological states so that they become analyzable on their own, not just their relationship with each other. With the inclusion of nonsignificantly changing data in the analysis, fewer apparent global and mesoscopic changes can be identified among the three networks. This work focuses on the structural reorganization of the network modules, for which it is inevitable to understand their biological function. In our model, modules mainly represent protein complexes. The biological role of a protein complex is determined by all members; therefore, the structural reorganization in itself, without any significant fold change, can be informative, as shown in this article.

2.2. Human Cancer Signaling Network. The network used for the colon cancer gene expression analysis was the Human Cancer Signaling Network constructed by Cui et al.[27] It was built after a comprehensive analysis of cancer signaling from genes that were frequently mutated in cancer based on, inter alia, the COSMIC,[31] BioCarta,[32] and Cancer Cell Map databases. This network contains 1634 nodes (that mainly represent proteins) and 5089 directed links that include 2403 activating links, 741 inhibiting links, 1915 undirected (association-type or interaction-based) links, and 30 links of unknown types. After mapping this network with the gene expression data set and deleting the interactions that were not covered by it or whose link type was not known and the nodes that were not connected to the giant component of the network, 1600 nodes and 5060 links remained, a network large enough not only for the analysis of local alterations but also for that of global topological changes during carcinogenesis.

2.3. Mapped Gene Expression Omnibus Data Sets. The healthy, adenomatous, and carcinomatous colon transcriptomic data for this study were extracted from the Gene Expression Omnibus database. The GSE20916,[33] GSE4183,[34] GSE8671,[35] GSE37364,[36] and GSE33113[37] series were processed, with the gene expression data of altogether 437 samples from patients of various ethnicities. Out of these samples, 128 were classified as healthy colons, 131 as adenomas, and 178 as colon adenocarcinomas. (See Table S1.) Data were collected with differentiation between left- and right-sided colon cancer. To avoid internal heterogeneity, the data sets selected contain mainly left-sided samples, which are known to be less immunogenic and richer in cancer-signaling-related mutations. The clinical stage in the colon cancer samples were diversely collected, aiming for a "general cancer phenotype".

2.4. Conversion of mRNA Expression Data to Link Weights. After collecting the data, RNA normalization and median aggregation were conducted for each of the mRNA data points. The median of every GEO[29,30] gene expression data set (which we refer to as abundance) in all three states (normal, adenoma, and carcinoma) was calculated respectively for each of the genes. While mapping the data to the structure of the Human Cancer Signaling Network,[27] the link weights were calculated by multiplying the abundances of the two nodes belonging to a given link:

$$\text{link weight}_{AB} = \log_{10}(\text{abundance}_A \times \text{abundance}_B)$$

The reason for this formula lies in the assumption that multiplication can highlight the upregulated genes in cancer. Additionally, because a significant part of the Human Cancer Signaling Network[27] is based on protein−protein interactions, in this case the probability of the attachment correlates with the multiplied abundances of the two interacting proteins.[38−40] For further calculations, the link weights were logarithmically transformed in order to avoid the distortions caused by some highly outlying gene expression values. The result was three sets of link weights representing the three data sets collected from normal colon, colon adenoma, and colon adenocarcinoma samples. The abstract models of these biological entities were defined in the text as a normal network for normal, healthy colons, an adenoma network for adenomatous colons, and a carcinoma network for cancerous colons.

In this study, mRNA measures were used as a proxy for protein abundance. The authors are aware that this is only an approximation, but the current lack of ample high-quality and extensive proteomic data justifies this estimation. Link weights (calculated as the product of node abundances) therefore characterize the probability of the given signaling interaction, which is high when the abundance of interacting nodes is high and low when it is low.

https://dx.doi.org/10.1021/acs.jpcb.0c09307
J. Phys. Chem. B 2021, 125, 1716−1726

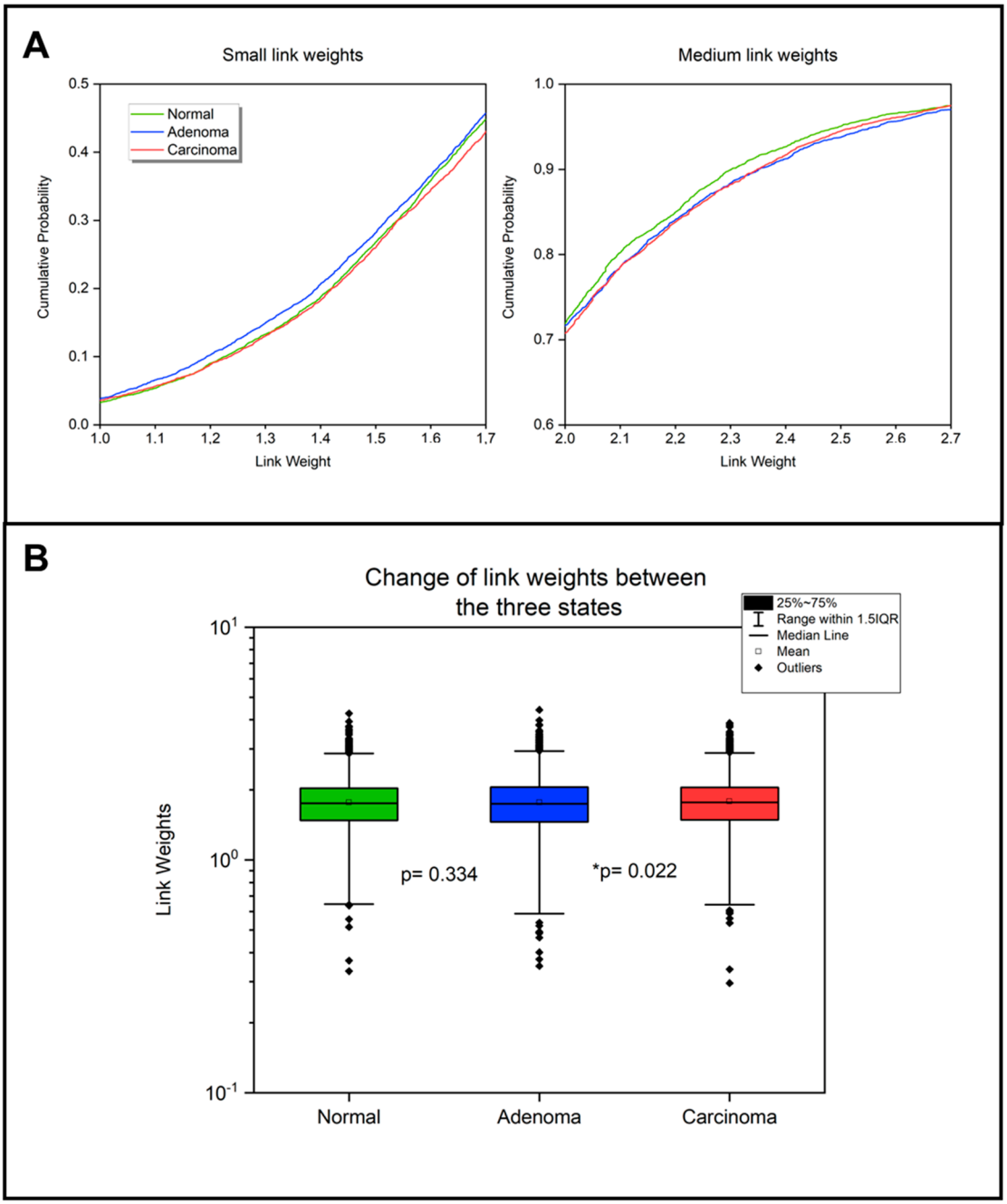


Figure 1. Link weight changes between the normal colon, colon adenoma, and colon carcinoma networks. The networks and link weights were created as described in the Materials and Methods section. (A) Cumulative probability distribution of the link weights in the three different states. In the two graphs, link weights from 1 to 1.7 and from 2 to 2.7 are highlighted. The first graph shows that link weights of the adenoma network have a higher cumulative probability, meaning that the number of smaller link weights (defined as a link weight of less than 2) is higher than in normal and carcinoma networks. On the second graph, the normal network's cumulative probability values exceed those of the others, which indicates that in the normal network the medium-weighted links (defined as a link weight of between 2 and 2.7) are more accentuated. (B) Box plots of the link weight data of the three different states. The *p* values between each two networks (normal−adenoma and adenoma−carcinoma; paired Wilcoxon test) are highlighted on the panel. These results together (along with the changes regarding the mean and median values; see the main text) suggest that the weakest links are becoming stronger in the carcinoma network, thus showing the realignment of the network in cancer.

While microarray data are known to be quite noisy,[33] signaling networks are known to be quite robust to noise.[9] To show the robustness of the results, all significant measurements were performed with artificially added noise by randomly increasing or decreasing the abundances by 5%. (For further details, see Text S1.)

2.5. Calculation of Network Diameter. For the three different networks, three different cases were considered: undirected, directed, and mixed networks. Undirected and directed cases contained only undirected and directed links, respectively. The mixed network contained both directed and undirected links, and the undirected links were substituted with two directed links of the same link weight. The calculation of the network diameter was conducted via the NetworkX Python package,[41] which is widely used for network analysis. Because NetworkX cannot calculate the network diameter of a

https://dx.doi.org/10.1021/acs.jpcb.0c09307
J. Phys. Chem. B 2021, 125, 1716−1726

non-singly connected component, removing the isolated parts was the next step in processing. Since NetworkX considers link weights to be path lengths between two nodes, transformation of the link weights was necessary for the appropriate representation of the probability of the biological connection between the nodes in the diameter calculation. As a next step, negative logarithmic mapping was performed so that the shortest distance mapped into the most likely transition path. The computation of the diameter was then executed with the help of the built-in Dijkstra algorithm[42] of NetworkX. The last step was choosing the longest of the calculated shortest paths and summing the link weights belonging to each of the links. Considering the fact that although the network diameter is traditionally defined as the longest of the shortest paths, in the case of weighted networks the average path length may be more reliable, which was also calculated. In summary, the treatment of the undirected link weights in the mixed and undirected networks may be considered to be an improvement in the diameter calculation. For further details, see Codes S1−S3.

2.6. EntOpt Layout Cytoscape Plugin. The EntOpt Layout program[43] is a network visualization plugin of the Cytoscape software environment for the analysis of biomolecular interaction networks. Its principle is relative entropy optimization.[44] In other words, it aims to arrange networks in a way that involves the least information loss (i.e., relative entropy), resulting in an easily interpretable and visually pleasing appearance. Entropy-based visualization also helps to highlight the global differences between the specific states in the networks, as the change in the link weights and the strengthening or weakening of different regions affect the visual output of the program. The EntOpt Layout is a unique tool for visualizing modules by grouping nodes together with a locally higher link ratio, also for weighted networks. Our aim was to visualize the pre-established modules (detected with the ModuLand plugin, see Section 2.7), where the nodes usually have similar biological functions. Therefore, with the use of the square adjacency matrix function, the nodes were put together on the basis of the neighborhood similarity, so nodes close to each other are not necessarily connected. We used the recommended visualization settings provided by the authors of the EntOpt Layout to display the Human Cancer Signaling Network[27] which are detailed in Text S2 and ref 43.

2.7. ModuLand Cytoscape Plugin. The ModuLand plugin[45] of Cytoscape is a Java-based network analyzing program focused on extensively overlapping modules and hierarchical layers (Figures S1−S3). This embeddedness and overlap in terms of the division of labor between functional units such as protein complexes is widely spread in living, real-world networks (e.g., cells). In this work, the authors used biological networks to model protein−protein interactions and signaling cascades of cells. For this reason, the ModuLand plugin was a useful tool for understanding the functional modular changes comparing the three networks (normal, adenoma, and carcinoma networks). It calculates the assignment strength of each node to each discovered module based on influence functions and centrality, creating a community landscape of overlapping hills (modules) of nodes. One of the main outputs of the plugin is the community centrality measure, which determines the sum of local influence zones for each node, also representing the whole network's influence on one of its nodes[28] Nodes having the largest community centrality measure in the module they mostly belong to are the main integrators and organizers of their module. On the basis of this property, estimating the biological function of modules by their leading nodes in their community centrality measure is a more precise way than checking each node in the module separately. The program also calculates several other topological measures, such as the modular overlap and effective degree.

3. RESULTS

3.1. Link Weight Changes. To better understand the global topological changes in the network during carcinogenesis, the differences in the link weight data (calculated as a product of node abundances, see Materials and Methods) were analyzed in the three states of the network. To emphasize the importance of the differences among the three data sets, the authors note that the structure of the network remains the same in normal, adenoma, and carcinoma states, and the diversity of the results stems only from the change in the mRNA expression data. For easier comprehension, the authors considered the normal network to be a model network of a normal, healthy colon, the adenoma network to be that of a colon adenoma, and the carcinoma network to be that of a model network of colon adenocarcinoma.

The distribution of the link weights (Figure 1A and Figure S4) shows a slightly different pattern in the three networks representing each biological state, which is further demonstrable with data binning (Figures S5 and S6). Among the weak links (defined here as link weights of less than 2), the adenoma network has the highest cumulative probability, demonstrating lower link weights (minimum is 0.0797) compared to the other states. Normal and carcinoma networks both have larger minimum link weights (0.333 and 0.296, respectively), indicating a similarity in normal and carcinoma networks. The normal network has the highest cumulative probability among the medium-strength links (link weights of between 2 and 2.7), thus having the lowest third quartile (2.035) and interquartile range (0.556). Strong links (link weights greater than 2.7) are very prominent in the adenoma network (maximum link weight of 4.414) and less expressed in the carcinoma network (maximum link weight of 3.87), while the normal network is in between the carcinoma and adenoma networks. Nonlogarithmic link weights and abundances also show these differences, while they are less accentuated with the weighted degree distribution (Figures S7−S9). These results are also robust to noise (Figures S10 and S11).

The differences that are found can also be characterized by the changes in the median, which is the smallest in the adenoma network (1.743), the largest in the carcinoma network (1.767), and in between in the normal network (1.7505), as visible in the box plot of the link weights of the three networks (Figure 1B). The box plot also demonstrates the standard deviations: in the adenoma network, it is the largest (0.4602) and much smaller in the normal and carcinoma networks (0.43349 and 0.44459, respectively). The differences among the link weight distributions of the three networks are also significant (paired Wilcoxon test; $p < 0.0001$), even with 5% added noise (paired Wilcoxon test; $p < 0.0001$). These small but consequent global differences highlight subtle changes among the three networks. In the next sections, with the use of more link-weight-sensitive methods, the functional implications of these changes have been explored. In conclusion, normal and carcinoma networks seem to have a very similar link weight distribution among the

https://dx.doi.org/10.1021/acs.jpcb.0c09307
J. Phys. Chem. B 2021, 125, 1716−1726

Table 1. Network Diameters in the Normal, Adenoma, and Carcinoma Networks with Negative Logarithmic Mapping

	undirected[a]		directed[b]		mixed graph[c]	
	network diameter[d]	average path length[e]	network diameter[d]	average path length[e]	network diameter[d]	average path length[e]
normal	34.361	9.984	36.988	12.940	36.435	11.397
adenoma	36.365	10.667	40.051	13.790	39.152	12.198
carcinoma	29.753	8.224	32.499	10.798	30.769	9.423

[a]In the network, every link was assigned as undirected. [b]The original directivity was preserved in the calculation. [c]Mixed graphs were defined as directed graphs, where undirected links were considered to be bidirectional links. [d]Network diameters were calculated with the Dijkstra algorithm. [e]Average path lengths were calculated with the NetworkX package.

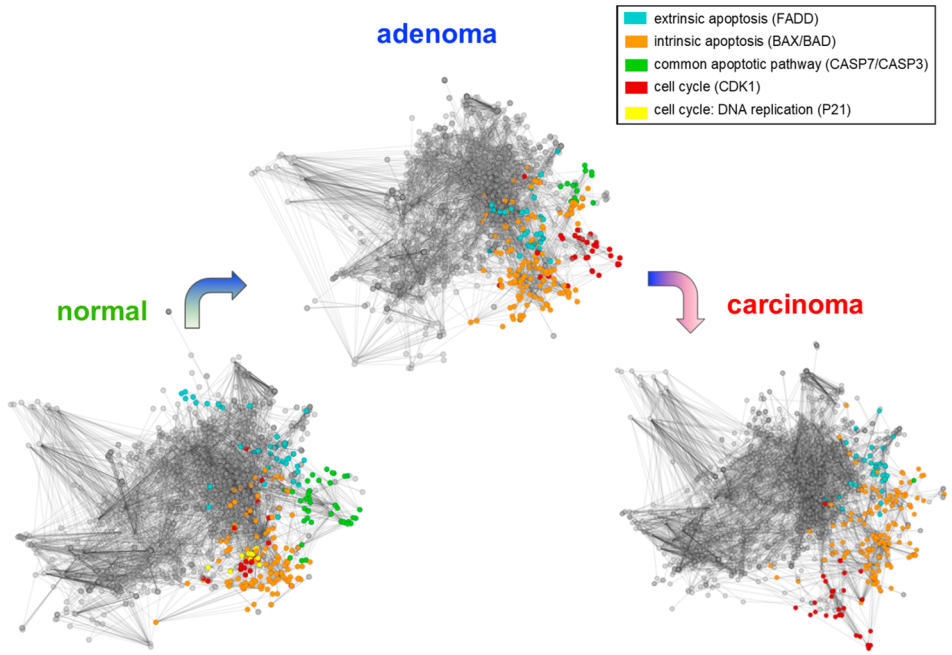


Figure 2. Functional changes in the modules of apoptosis and the cell cycle. The Human Cancer Signaling network was visualized using the EntOpt Cytoscape plugin.[40] The most relevant modular changes were chosen and colored. The functions of the modules were estimated from the community centrality measure of the ModuLand plugin[45] (Materials and Methods). The module responsible for the common apoptotic pathway (named CASP3 in normal and CASP7 in adenoma networks) is continuously melting into the intrinsic apoptotic pathway module (named BAD in the normal network and BAX in adenoma and carcinoma networks), while the structure and size of the extrinsic apoptotic module remain stable (named FADD). The module responsible for controlling the cell cycle and DNA replication (named P21) melts into the CDK1 module. A small part remains separate, named the PCNA module. (For details, see Table S3.) These functional changes suggest that the control of the apoptotic process and the cell cycle becomes less organized and a less important part of the network, in line with our knowledge of tumor biology. These results showed a robustness to noise (Table S4).

small and large link weights, but among the medium link weights, the normal network becomes the most prominent.

3.2. Network Diameter. The network diameter by definition is the sum of the link weights within the longest shortest path between two nodes.[46] However, in order to calculate this, a nonconventional approach was necessary. In the processed molecular networks, weights are considered to be the likelihood of the interaction between two nodes (which represent molecules in this case); therefore, the higher the weight value, the shorter the path, while in graph theory, weights are considered to be lengths. To overcome this issue, during the calculation of the longest shortest path, the link weights were negated, and after running the Dijkstra algorithm,[42] the weighted network diameter was calculated by summing the negatively mapped link weights in the longest shortest path of the network (further details in Materials and Methods). For further evaluation of the results, the average shortest path length was calculated with the same method.

Link directivity raised another complex methodical question concerning whether the direction of the connections should be included in the calculation. Ultimately, three methods were used. First, the network was considered to be completely undirected, and then the directivity of activating and inhibiting links was respected but not that of undirected links. Finally, a mixed graph was created by duplicating undirected links by taking them as two different connections pointing in the opposite direction (Table 1).

Out of all three networks, the adenoma network has the largest and the carcinoma network has the smallest network diameter by all calculation methods. The average shortest path length changes accordingly. The same results were received with two different approaches and also with additional noise (Table S2). The enlarged network diameter in the adenoma network may imply a less compact network structure. The shrinkage of the network diameter and average shortest path length in the carcinoma network may be an adaptation strategy by making the information flow in the network more effective, which is already observed in other topological studies. This result may represent the cellular response to stress in a

https://dx.doi.org/10.1021/acs.jpcb.0c09307
J. Phys. Chem. B 2021, 125, 1716−1726

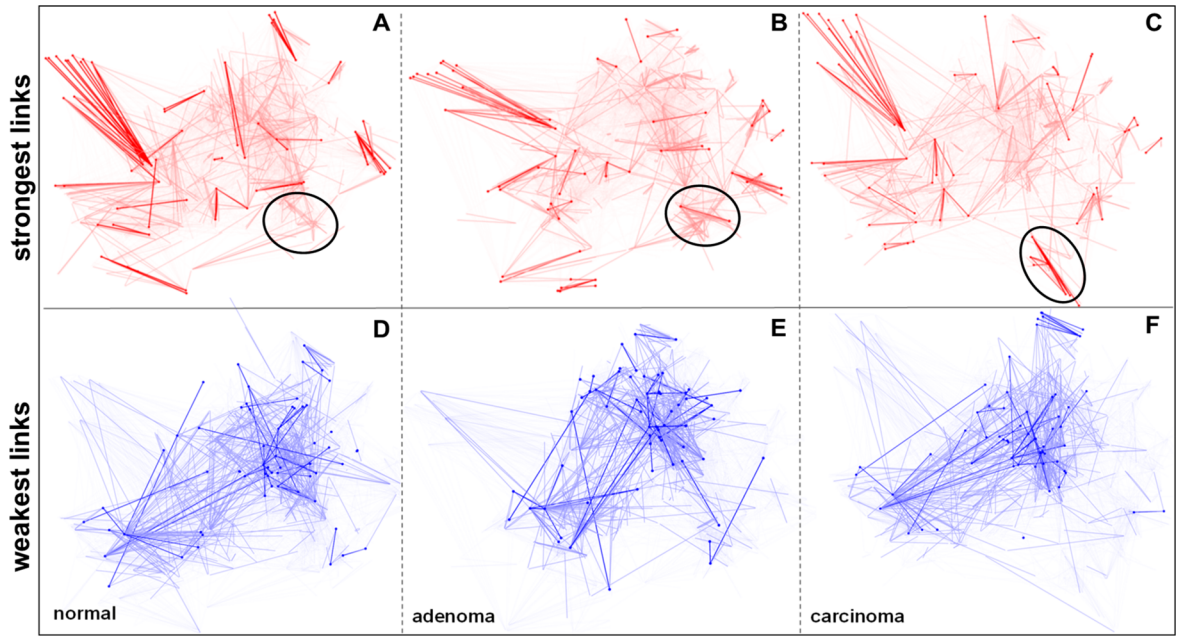


Figure 3. Strongest and weakest links of normal (A, D), adenoma (B, E), and carcinoma (C, F) colon data applied to the Human Cancer Signaling Network. The network was visualized using the EntOpt Cytoscape plugin.[43] For more details, see Tables S5 and S6 (panels A−C). The top 1% of the strongest links are highlighted in red, while the brighter red lines show the top 2 to 10%. The rounded part shows the most relevant strengthening area in the adenoma and carcinoma networks, banding together as the CDK1 module shown in Figure 3 (panels D−F). The bottom 1% of the weakest links are highlighted with blue, while the brighter blue lines show the bottom 2 to 10%.

cancerous environment, offering new possibilities for interpretation compared to changes during carcinogenesis.

3.3. Functional Changes in the Modules. With the discrete module assignment method, the ModuLand plugin assigns the nodes to the module to which they mostly belong.[45] The name of the module often depicts its function, but for a more precise estimation of the biological functions, the nodes with the highest community centrality measure within the module were analyzed. Community centrality refers to the local influence zone of each node and thus helps in the determination of a given node's dominance within its neighborhood.[28] The nodes chosen with the above-described approach were checked in the UniProt database[47] (in which each node refers to a protein in the network), and the summarized biological functions were set to be the functions of the modules.

The most important modular changes in the network were in modules associated with cell cycle and apoptosis (Figure 2). The appropriate regulation of these functions is essential to the survival of cancer cells. One of the most important network changes in the process of carcinogenesis is that the module responsible for the final common pathway of apoptosis (called CASP3 in the normal network and CASP7 in the adenoma network) gradually merges with the module responsible for intrinsic apoptosis (called BAX), suggesting a decrease in the importance of apoptotic functions. The p21 protein is responsible for cell cycle arrest by inhibiting cyclin-dependent kinases, also acting as an effector in the DNA damage-induced p53-mediated apoptosis in colon adenocarcinoma.[48] Thus, the fusion of the module organized around p21 in the CDK1 module in the adenoma and carcinoma networks indicates the loss of function of cell cycle control during these pathologic transitions.

Interestingly, the EntOpt images of normal and carcinoma networks are very similar to each other, and the adenoma network displays a slightly more compact network periphery compared to the other two networks (Figures S12−S15). This observation may correlate with the findings that several topological parameters of the carcinoma network are closer to those of the normal network than to the topological parameters of the adenoma network.

There was a further investigation into the change in the modular overlap among the three networks, but this analysis did not bring about significant differences. (For further details, see Text S3 and Figures S16−S19.)

3.4. Strongest and Weakest Links in the Network. The process of carcinogenesis entails the strengthening and weakening of links according to the underlying changes in the biological functions of the cell. The distribution of the link weights (as seen in Figure 1) provides information about the general differences between the networks highlighting the strongest and weakest links (by picking the first 10% of the highest and lowest link weight values, further separating the top 1% into a subcategory Tables S5 and S6) and focusing on their functional changes adds another point of view to this work. Among the strongest links, the most prominent change is the strengthening of the cell cycle-associated links belonging to the CDK1 module (Figure 3A−C), and this result is also robust to noise (Table S7). This is consistent with the observed modular changes in the network and is associated with the inspected constant proliferation in cancer cells. Other permanently highly expressed parts of the network are responsible for immunology-associated functions such as motility and the cytokine response. The weak links contain vegetative regulation- and neuronal transmission-associated links, which make up a remarkable part of the utilized Human Cancer Signaling Network, but in colon cells, these functions are not really highlighted (Figure 3D−F). For the same reason, the structure of these links does not change among the different states of the network. In conclusion, strong and weak

https://dx.doi.org/10.1021/acs.jpcb.0c09307
J. Phys. Chem. B 2021, 125, 1716−1726

Table 2. Analysis of Targeted and Immunotherapy Pathways

	no. of associated proteins in gene ontology[a]	n umber of proteins in the network[b]	median normal abundance[c]	median adenoma abundance[c]	median carcinoma abundance[c]	p value (normal-adenoma)[d]	p value (adenoma-carcinoma)[d]	p value (normal-carcinoma)[d]
EGFR signaling	121	58	7.061	6.665	7.058	0.01783*	0.018*	0.67023
VEGFR signaling	93	49	6.662	7.383	7.608	0.84231	<0.001*	0.01791*
mismatch repair	38	8	8.533	9.931	8.987	0.08006	1.000	0.23395

[a]Number of human protein search results for GO terms: epidermal growth factor receptor signaling pathway, vascular endothelial growth factor receptor signaling pathway, and mismatch repair. [b]Number of proteins in the Human Cancer Signaling Network, constructed as described in the Materials and Methods. [c]Abundances were calculated from GEO data sets. For further details, see the Materials and Methods. We used median values because the data set distribution, based on Shapiro-Wilk tests, were not considered a normal distribution. [d]On the basis of the results of the normality tests, we used paired Wilcoxon tests to evaluate the significance of the changes in the targeted therapy-associated data.

link structure can promote important topological changes during carcinogenesis, and the growing importance of cell cycle regulation supports the raison d'être of cancer network analysis. Link weights proved to be more sensitive to the detection of biological changes in the network than the mRNA abundances and weighted degrees, after a comparison of the representation ratio of the cell cycle and apoptosis-related modules (Table S8).

3.5. Changes in Targeted Therapy and Immune Checkpoint Inhibitor Associated Pathways. Targeted therapies represent a new and innovative approach in oncology. By the targeted inhibition of different signaling pathways, survival and the quality of life of metastatic colon cancer patients can be largely increased. In this research, the signaling pathways of two widely used therapeutic approaches in colon cancer, EGFR and VEGFR inhibition, and the effectiveness of immune checkpoint inhibition are estimated in advance to facilitate therapeutic decision making by examining microsatellite instability, and the loss of expression of the mismatch repair proteins responsible for it was examined by analyzing the changes in the gene expression profile of the proteins associated with the appropriate Gene Ontology terms. In addition, the effectiveness of immune checkpoint inhibition is estimated in advance to facilitate therapeutic decision making by examining microsatellite instability and the loss of expression of the mismatch repair proteins responsible for it. The mRNA expression of these proteins was also analyzed in this study by using the relevant gene ontology term (Text S4 and Table S9). Eventually, about half of the proteins associated with EGFR and VEGFR signaling and only a quarter of the proteins in the mismatch repair pathway were found in our network. The reason for this could be that the Human Cancer Signaling Network mainly contains proteins associated with signaling, thus not many proteins in the mismatch repair pathway are included.

To examining the differences among normal, adenoma, and carcinoma networks, the median abundance of each of the pathways was calculated, and for the verification of the significance of the results, paired Wilcoxon tests were conducted, as the data could have not been considered to be normally distributed (Materials and Methods and Table 2). The EGF receptor signaling was shown to be very important in the normal network, which may be because the EGF is known to be an important mediator in the alimentary tract.[49] In the adenoma network, there is a significant reduction in its importance, showing that it is not the EGFR signaling that drives this pathologic transition.[50] In the carcinoma network, the median abundance is almost the same as in the normal network, indicating a significant difference between the adenoma and carcinoma networks. The decreasing EGF receptor mRNA expression in our data (8.75 in the normal network, 8.06 in the adenoma network, and 5.62 in the carcinoma network) may be caused by the possibly remarkable ratio of BRAF mutant transcriptomic data in our data set (Materials and Methods), when the EGFR signaling pathway is constantly activated by the mutant BRAF protein.[51] The median abundance of the VEGF receptor signaling is constantly growing among the three networks, which is significant between adenoma and carcinoma and between normal and carcinoma networks. It may be explained by that the level of VEGF is well known to be associated with cellular hypoxia, which is an important attribute of the cancer microenvironment.[52] The proteins associated with mismatch repair have the largest median abundance in the adenoma network, and it is slightly decreasing between adenoma and carcinoma networks. However, these changes are not significant, which may be explained by the selection criteria for transcriptomic data, as it mainly consists of left-sided colon cancer (further details in Materials and Methods).

For further characterization of the targeted- and immunotherapy-associated nodes, median weighted degrees were calculated for each pathway. In the case of VEGFR and EGFR pathways, the median weighted degrees were more than twice those of the whole networks but not of the mismatch repair-related nodes (Table S10). The modular affiliation of each node in each network was also determined. The majority of the nodes belong to the RAC1 module, which is the largest module of the network, with non-definitively ascertainable biological functions. As a consequence, these pathways do not form distinctive modules in the network probably because they are so intertwined with every part of the graph that they cannot be separated by the modularization program (Figures S20–S22).

Links of the targeted- and immunotherapy-associated nodes can be found among the strongest and weakest links in the network with altering dynamics among the three states. The number of EGFR signaling-associated links among the top and bottom 10% of the links in the order of link weights is decreasing for the normal network (top, 42; bottom, 38), adenoma network (top, 27; bottom, 36), and carcinoma network (top, 26; bottom, 32). In VEGFR signaling, the number of links in the top 10% of the links is 63 in the normal network, 52 in the adenoma network, and 73 in the carcinoma network. The number of links in the bottom 10% of the links is steadily decreasing. In terms of the mismatch repair, there are not enough proteins in the network to see a big difference, but

https://dx.doi.org/10.1021/acs.jpcb.0c09307
J. Phys. Chem. B 2021, 125, 1716−1726

the decreasing expression of the pathway between the adenoma and carcinoma networks may indicate increased microsatellite instability. In conclusion, targeted therapy-associated pathways gain importance in the carcinoma network.

4. DISCUSSION

The aim of this work was to analyze the changes in the network topological characteristics during the development of colon adenoma and adenocarcinoma. In the adenoma network, the strongest link weights get stronger and the weakest link weights get weaker, causing an increase in the standard deviation and implying the importance of a few highly expressed proteins. In the carcinoma network, the median link weights are larger than in the normal network; however, both the standard deviation and maximum link weights are decreasing, implying a more compact link weight distribution. Network diameter analysis also shows structural differences among the adenoma and carcinoma networks. The network diameter is the largest in the adenoma network and the smallest in the carcinoma network. This may indicate a strengthened shortcut system in cancer cells, which may help them to evade drug treatment. Modular reorganization and the analysis of the strongest and weakest links support the emergence of cancer hallmarks of constant proliferation and evading apoptosis in the carcinoma network since the cell cycle comes into prominence and the apoptotic modules unite. The steady strengthening of VEGFR and the relative strengthening of EGFR pathways compared to the adenoma network are particularly important results because the relevance of these pathways is clinically proven as these signaling pathways are involved in the treatment of advanced colon cancer.

The decrease in network diameter between normal and carcinoma networks was described multiple times in other previous topological studies.[10−15] Our results showing the decrease in the network diameter between normal and carcinoma networks agree with those of several formal studies conducted on differential coexpression,[10,12] association-type,[11] protein−protein interaction,[14,15] and ceRNA networks[13] but not on signaling networks. The study by Wen et al. used a method very similar to that used in this study but with miRNA networks;[14] namely, they mapped expression data on a pre-existing literature-based network structure. One particular study that found differences in the change in network diameter from normal to carcinoma networks between multiple types of cancer describe colon cancer as a type of cancer with decreasing diameter.[53] The other results indicating high levels of compactness are unprecedented in the literature and may be part of a new approach to the description of the development of colon cancer at the system level. The strengthening of the cell cycle and the downregulation of apoptosis in cancer networks were also determined via multiple methods in previous studies.[17−20] According to the global topological results, the carcinoma network seems to be closer to the normal network in the aspect of topological parameters than to the adenoma network, but with regards to the mesoscopic network parameters such as modular properties and the strongest and weakest links, signs of steady progression are detected.

The relevance of the targeted therapy pathways is apparent from the clinical trials, showing the efficacy of these treatments. The steadily increasing median abundance of the VEGFR pathway in this research is in line with the efficacy of VEGFR inhibitors, such as bevacizumab[54] and ramucirumab[55] in metastatic colon cancer. The similar median abundance of the EGFR pathway in the normal and the carcinoma networks requires further evaluation, but the EGF is already known to play an essential role in healthy alimentary tracts.[50] The strengthening of the EGF receptor signaling pathway between colon adenoma and carcinoma was previously implied in the literature, as an increase in the EGFR copy number is detected[51] and EGFR inhibitors such as cetuximab[56,57] and panitumumab[58] are widely used in therapy. The results with the targeted and immunotherapy-related pathways are important methodological feedback. Arriving at the same conclusions with network science as with molecular and clinical studies confirms the relevance of the methodology used in this study.

Our work paves the way for further research into different types of cancer networks and demands proof of the generality of systemic changes in carcinogenesis, such as regarding the link weight distribution and network diameter. The individual differences between cancer types which drive carcinogenesis and determine therapeutic possibilities can be assessed by exploring modular rearrangement and pathway enrichment. Precision oncotherapy is an emerging topic in oncology, and the analysis of gene expression data may play an important role in estimating the efficacy of different therapeutic choices.

5. CONCLUSIONS

The most relevant results of this work were that (i) the distribution of the link weights was spread by increasing the standard deviation in adenoma compared to those of the normal and carcinoma networks; (ii) the size of the weighted diameter is explicitly the largest in the adenoma network and the smallest in the carcinoma network; (iii) the modules of the apoptosis and cell cycle become reorganized in the adenoma and carcinoma networks; (iv) the strengthening of a group of links was observed in the area of the cell cycle in carcinoma; and (v) the observations regarding the pathways associated with targeted therapies are in good agreement with the current clinical knowledge. Our major, link-weight-based results are robust to noise. This work could provide a novel approach to cancer data analysis through modeling the development of colon cancer on the system level.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.0c09307.

Text S1. Check of the robustness of the results by adding noise to the data; Text S2. algorithm for the generation of EntOpt images; Text S3. investigation of the modular overlap changes; Text S4. method for the examination of targeted and immunotherapy-related pathways; Fig. S1.hierarchy of modules -- normal; Fig. S2. hierarchy of modules -- adenoma; Fig. S3. hierarchy of modules -- carcinoma; Fig. S4. distribution of logarithmic link weights; Fig. S5. number of links in different bins of data; Fig. S6. probability density function of the logarithmic link weights; Fig. S7. cumulative distribution of the nonlogarithmic link weights; Fig. S8. distribution of protein abundances; Fig. S9. distribution of weighted degrees; Fig. S10. distribution of link weights with noise; Fig. S11. box plot

https://dx.doi.org/10.1021/acs.jpcb.0c09307
J. Phys. Chem. B 2021, 125, 1716−1726

of the nonlogarithmic link weights with noise; Fig. S12. EntOpt images of the unweighted Human Cancer Signaling Network; Fig. S13. EntOpt image of the network with normal weights; Fig. S14. EntOpt image of the network with adenoma weights; Fig. S15. EntOpt image of the network with carcinoma weights; Fig. S16. changes in the ModuLand overlap values based on the different number of modules with logarithmic link weights; Fig. S17. changes in the ModuLand overlap values based on the different number of modules with nonlogarithmic link weights; Fig. S18. cumulative distribution of the effective degree of modules; Fig. S19. cumulative distribution of the normalized modular overlap; Fig. S20. EGFR and VEGFR signaling and mismatch repair-related nodes in the normal network; Fig. S21. EGFR- VEGFR-signaling and mismatch repair nodes -- the adenoma network; Fig. S22. EGFR-VEGFR-signaling and mismatch repair nodes -- carcinoma network; Table S1. number of samples in the data set series; Table S2. network diameter with additional 5% noise; Table S3. function of the largest network module nodes; Table S4. largest modules of the networks with additional 5% noise; Tables S5 and S6. relevant changes in the strongest and weakest 1% of the links and Table S7. in the strongest and weakest 1% of the links with additional 5% noise; Table S8. representation of the apoptosis- and cell cycle-related modules among the strongest and weakest 10% of the abundances; weighted degrees and the link weights; Table S9. nodes in the targeted and immunotherapy-related pathways; Table S10. median weighted degrees of targeted and immunotherapy pathway-related nodes; Table S11. calculation of the diameter for the undirected network, Table S12. the directed network, and Table S13. the mixed network (PDF)

AUTHOR INFORMATION

Corresponding Author

Peter Csermely − *Department of Molecular Biology, Semmelweis University, Budapest 1085, Hungary;* orcid.org/0000-0001-9234-0659; Phone: +36-1-459-1500; Email: csermely.peter@med.semmelweis-univ.hu

Authors

Klára Schulc − *Department of Molecular Biology, Semmelweis University, Budapest 1085, Hungary*

Zsolt T. Nagy − *Department of Molecular Biology, Semmelweis University, Budapest 1085, Hungary*

Sebestyén Kamp − *Turbine Ltd, Budapest, Hungary*

János Molnár − *Turbine Ltd, Budapest, Hungary*

Daniel V. Veres − *Department of Molecular Biology, Semmelweis University, Budapest 1085, Hungary; Turbine Ltd, Budapest, Hungary*

Borbála M. Kovács − *Department of Molecular Biology, Semmelweis University, Budapest 1085, Hungary*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.0c09307

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank the anonymous referee of this article for excellent suggestions to better our manuscript. This work was supported by the Hungarian National Research, Development and Innovation Office (K131458), the Higher Education Institutional Excellence Programme of the Ministry of Human Capacities in Hungary within the framework of the molecular biology thematic programmes of Semmelweis University, the Thematic Excellence Programme (Tématerületi Kiválósági Program, 2020-4.1.1.-TKP2020) of the Ministry for Innovation and Technology in Hungary within the framework of the molecular biology thematic programme of the Semmelweis University, and the ÚNKP-19-2-I-SE-54 New National Excellence Program of the Ministry of Human Capacities. The LINK network science group (http://linkgroup.hu) where this work was also born gratefully acknowledges the guidance and support of Prof. Ruth Nussinov, both scientifically and personally, for the last 10 years.

REFERENCES

(1) Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R. L.; Torre, L. A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca-Cancer J. Clin.* **2018**, *68* (6), 394−424.

(2) Grady, W. M.; Markowitz, S. D. The molecular pathogenesis of colorectal cancer and its potential application to colorectal cancer screening. *Dig. Dis. Sci.* **2015**, *60* (3), 762−72.

(3) Fearon, E. R.; Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **1990**, *61* (5), 759−67.

(4) Bardhan, K.; Liu, K. Epigenetics and colorectal cancer pathogenesis. *Cancers* **2013**, *5* (2), 676−713.

(5) Ciombor, K. K.; Bekaii-Saab, T. A Comprehensive Review of Sequencing and Combination Strategies of Targeted Agents in Metastatic Colorectal Cancer. *Oncologist* **2018**, *23* (1), 25−34.

(6) Mendik, P.; Dobronyi, L.; Hari, F.; Kerepesi, C.; Maia-Moco, L.; Buszlai, D.; Csermely, P.; Veres, D. V. Translocatome: a novel resource for the analysis of protein translocation between cellular organelles. *Nucleic Acids Res.* **2019**, *47* (D1), D495−D505.

(7) Schadt, E. E.; Bjorkegren, J. L. NEW: network-enabled wisdom in biology, medicine, and health care. *Sci. Transl. Med.* **2012**, *4* (115), 115rv1.

(8) Gosak, M.; Markovic, R.; Dolensek, J.; Slak Rupnik, M.; Marhl, M.; Stozer, A.; Perc, M. Network science of biological systems at different scales: A review. *Phys. Life Rev.* **2018**, *24*, 118−135.

(9) Csermely, P.; Korcsmaros, T.; Kiss, H. J.; London, G.; Nussinov, R. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol. Ther.* **2013**, *138* (3), 333−408.

(10) Deng, S. P.; Zhu, L.; Huang, D. S. Predicting Hub Genes Associated with Cervical Cancer through Gene Co-Expression Networks. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2016**, *13* (1), 27−35.

(11) Lee, Y. S.; Hwang, S. G.; Kim, J. K.; Park, T. H.; Kim, Y. R.; Myeong, H. S.; Kwon, K.; Jang, C. S.; Noh, Y. H.; Kim, S. Y. Topological network analysis of differentially expressed genes in cancer cells with acquired gefitinib resistance. *Cancer Genomics Proteomics* **2015**, *12* (3), 153−166.

(12) Kugler, K. G.; Mueller, L. A.; Graber, A.; Dehmer, M. Integrative network biology: graph prototyping for co-expression cancer networks. *PLoS One* **2011**, *6* (7), No. e22843.

(13) Zhou, M.; Wang, X.; Shi, H.; Cheng, L.; Wang, Z.; Zhao, H.; Yang, L.; Sun, J. Characterization of long non-coding RNA-associated ceRNA network to reveal potential prognostic lncRNA biomarkers in human ovarian cancer. *Oncotarget* **2016**, *7* (11), 12598−611.

(14) Wen, J.; Hall, B.; Shi, X. A network view of microRNA and gene interactions in different pathological stages of colon cancer. *BMC Med. Genomics* **2019**, *12* (Suppl 7), 158.

(15) Verma, Y.; Yadav, A.; Katara, P. Mining of cancer core-genes and their protein interactome using expression profiling based PPI network approach. *Gene Rep* **2020**, *18*, 100583.
(16) Yang, J.; Leskovec, J. Overlapping Communities Explain Core-Periphery Organization of Networks. *Proc. IEEE* **2014**, *102* (12), 1892−1902.
(17) Qi, Y. W.; Qi, H. W.; Liu, Z. Y.; He, P. Y.; Li, B. Q. Bioinformatics Analysis of Key Genes and Pathways in Colorectal Cancer. *J. Comput. Biol.* **2019**, *26* (4), 364−375.
(18) Wen, Z. S.; Liu, Z. P.; Liu, Z. R.; Zhang, Y.; Chen, L. N. An integrated approach to identify causal network modules of complex diseases with application to colorectal cancer. *J. Am. Med. Inform Assn* **2013**, *20* (4), 659−667.
(19) Liu, R.; Zhang, W.; Liu, Z. Q.; Zhou, H. H. Associating transcriptional modules with colon cancer survival through weighted gene co-expression network analysis. *BMC Genomics* **2017**, *18*, 361.
(20) Qu, X. L.; Xie, R. Q.; Chen, L. N.; Feng, C. C.; Zhou, Y. Y.; Li, W.; Huang, H.; Jia, X.; Lv, J. J.; He, Y. H.; et al. Identifying colon cancer risk modules with better classification performance based on human signaling network. *Genomics* **2014**, *104* (4), 242−248.
(21) Zhou, X. G.; Huang, X. L.; Liang, S. Y.; Tang, S. M.; Wu, S. K.; Huang, T. T.; Mo, Z. N.; Wang, Q. Y. Identifying miRNA and gene modules of colon cancer associated with pathological stage by weighted gene co-expression network analysis. *OncoTargets Ther.* **2018**, *11*, 2815−2830.
(22) Yu, H. L.; Ye, L.; Wang, J. X.; Jin, L.; Lv, Y. F.; Yu, M. Protein-protein interaction networks and modules analysis for colorectal cancer and serrated adenocarcinoma. *J. Cancer Res. Ther* **2015**, *11* (4), 846−851.
(23) Petrochilos, D.; Shojaie, A.; Gennari, J.; Abernethy, N. Using random walks to identify cancer-associated modules in expression data. *BioData Min.* **2013**, *6*, 17.
(24) Csermely, P.; Korcsmaros, T.; Nussinov, R. Intracellular and intercellular signaling networks in cancer initiation, development and precision anti-cancer therapy RAS acts as contextual signaling hub. *Semin. Cell Dev. Biol.* **2016**, *58*, 55−59.
(25) Nussinov, R.; Tsai, C. J.; Jang, H.; Korcsmaros, T.; Csermely, P. Oncogenic KRAS signaling and YAP1/beta-catenin: Similar cell cycle control in tumor initiation. *Semin. Cell Dev. Biol.* **2016**, *58*, 79−85.
(26) Nussinov, R.; Tsai, C. J.; Csermely, P. Allo-network drugs: harnessing allostery in cellular networks. *Trends Pharmacol. Sci.* **2011**, *32* (12), 686−93.
(27) Cui, Q.; Ma, Y.; Jaramillo, M.; Bari, H.; Awan, A.; Yang, S.; Zhang, S.; Liu, L.; Lu, M.; O'Connor-McCourt, M.; et al. A map of human cancer signaling. *Mol. Syst. Biol.* **2007**, *3*, 152.
(28) Kovacs, I. A.; Palotai, R.; Szalay, M. S.; Csermely, P. Community landscapes: an integrative approach to determine overlapping network module hierarchy, identify key nodes and predict network dynamics. *PLoS One* **2010**, *5* (9), No. e12528.
(29) Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25* (1), 25−9.
(30) Carbon, S.; Douglass, E.; Dunn, N.; Good, B.; Harris, N. L.; Lewis, S. E.; Mungall, C. J.; Basu, S.; Chisholm, R. L.; Dodson, R. J.; et al. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **2019**, *47* (D1), D330−D338.
(31) Tate, J. G.; Bamford, S.; Jubb, H. C.; Sondka, Z.; Beare, D. M.; Bindal, N.; Boutselakis, H.; Cole, C. G.; Creatore, C.; Dawson, E.; et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **2019**, *47* (D1), D941−D947.
(32) Nishimura, D. BioCarta. *Biotech Software & Internet Report: The Computer Software Journal for Scient* 2001, *2* (3), 117−120.
(33) Skrzypczak, M.; Goryca, K.; Rubel, T.; Paziewska, A.; Mikula, M.; Jarosz, D.; Pachlewski, J.; Oledzki, J.; Ostrowski, J. Modeling oncogenic signaling in colon tumors by multidirectional analyses of microarray data directed for maximization of analytical reliability. *PLoS One* **2010**, *5* (10), No. e13091.
(34) Galamb, O.; Gyorffy, B.; Sipos, F.; Spisak, S.; Nemeth, A. M.; Miheller, P.; Tulassay, Z.; Dinya, E.; Molnar, B. Inflammation, adenoma and cancer: objective classification of colon biopsy specimens with gene expression signature. *Dis. Markers* **2008**, *25* (1), 586721.
(35) Sabates-Bellver, J.; Van der Flier, L. G.; de Palo, M.; Cattaneo, E.; Maake, C.; Rehrauer, H.; Laczko, E.; Kurowski, M. A.; Bujnicki, J. M.; Menigatti, M.; et al. Transcriptome profile of human colorectal adenomas. *Mol. Cancer Res.* **2007**, *5* (12), 1263−75.
(36) Valcz, G.; Patai, A. V.; Kalmar, A.; Peterfia, B.; Furi, I.; Wichmann, B.; Muzes, G.; Sipos, F.; Krenacs, T.; Mihaly, E.; et al. Myofibroblast-Derived SFRP1 as Potential Inhibitor of Colorectal Carcinoma Field Effect. *PLoS One* **2014**, *9* (11), No. e106143.
(37) de Sousa E Melo, F.; Colak, S.; Buikhuisen, J.; Koster, J.; Cameron, K.; de Jong, J. H.; Tuynman, J. B.; Prasetyanti, P. R.; Fessler, E.; van den Bergh, S. P.; Rodermond, H.; Dekker, E.; van der Loos, C. M.; Pals, S. T.; van de Vijver, M. J.; Versteeg, R.; Richel, D. J.; Vermeulen, L.; Medema, J. P.; et al. Methylation of cancer-stem-cell-associated Wnt target genes predicts poor prognosis in colorectal cancer patients. *Cell Stem Cell* **2011**, *9* (5), 476−485.
(38) Xing, S.; Wallmeroth, N.; Berendzen, K. W.; Grefen, C. Techniques for the Analysis of Protein-protein Interactions in Vivo. *Plant Physiol.* **2016**, *171* (2), 727−758.
(39) Edwards, P. R.; Gill, A.; Pollardknight, D. V.; Hoare, M.; Buckle, P. E.; Lowe, P. A.; Leatherbarrow, R. J. Kinetics of Protein-protein Interactions at the Surface of an Optical Biosensor. *Anal. Biochem.* **1995**, *231* (1), 210−217.
(40) Mihalik, Á.; Csermely, P. Heat Shock Partially Dissociates the Overlapping Modules of the Yeast Protein-protein Interaction Network: A Systems Level Model of Adaptation. *PLoS Comput. Biol.* **2011**, *7* (10), No. e1002187.
(41) Hagberg, A.; Swart, P.; Chult, D. Exploring network structure, dynamics, and function using NetworkX. In *Proc. 7th Python Science Conference (SciPy 2008)* Varoquaux, G., Vaught, T., Millman, J., Eds.; Los Alamos National Laboratory (LANL): Los Alamos, NM, 2008; pp 11−15.
(42) Dijkstra, E. W. A note on two problems in connexion with graphs. *Numerische mathematik* **1959**, *1* (1), 269−271.
(43) Ágg, B.; Császár, A.; Szalay-Bekő, M.; Veres, D. V.; Mizsei, R.; Ferdinandy, P.; Csermely, P.; Kovács, I. A. The EntOptLayout Cytoscape plug-in for the efficient visualization of major protein complexes in protein−protein interaction and signalling networks. *Bioinformatics* **2019**, *35* (21), 4490−4492.
(44) Kovacs, I. A.; Mizsei, R.; Csermely, P. A unified data representation theory for network visualization, ordering and coarse-graining. *Sci. Rep.* **2015**, *5*, 13786.
(45) Szalay-Beko, M.; Palotai, R.; Szappanos, B.; Kovacs, I. A.; Papp, B.; Csermely, P. ModuLand plug-in for Cytoscape: determination of hierarchical layers of overlapping network modules and community centrality. *Bioinformatics* **2012**, *28* (16), 2202−2204.
(46) Amini, H.; Lelarge, M. The diameter of weighted random graphs. *Annals of Applied Probability* **2015**, *25* (3), 1686−1727.
(47) Bateman, A.; Martin, M. J.; Orchard, S.; Magrane, M.; Alpi, E.; Bely, B.; Bingley, M.; Britto, R.; Bursteinas, B.; et al. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47* (D1), D506−D515.
(48) Waldman, T.; Kinzler, K. W.; Vogelstein, B. p21 is necessary for the p53-mediated G1 arrest in human cancer cells. *Cancer Res.* **1995**, *55* (22), 5187−5190.
(49) Tang, X.; Liu, H.; Yang, S.; Li, Z.; Zhong, J.; Fang, R. Epidermal Growth Factor and Intestinal Barrier Function. *Mediators Inflammation* **2016**, *2016*, 1927348.
(50) Flora, M.; Piana, S.; Bassano, C.; Bisagni, A.; De Marco, L.; Ciarrocchi, A.; Tagliavini, E.; Gardini, G.; Tamagnini, I.; Banzi, C.; et al. Epidermal growth factor receptor (EGFR) gene copy number in colorectal adenoma-carcinoma progression. *Cancer Genet.* **2012**, *205* (12), 630−5.
(51) Corcoran, R. B.; Ebi, H.; Turke, A. B.; Coffee, E. M.; Nishino, M.; Cogdill, A. P.; Brown, R. D.; Della Pelle, P.; Dias-Santagata, D.;

https://dx.doi.org/10.1021/acs.jpcb.0c09307
J. Phys. Chem. B 2021, 125, 1716−1726

Hung, K. E.; et al. EGFR-Mediated Reactivation of MAPK Signaling Contributes to Insensitivity of BRAF-Mutant Colorectal Cancers to RAF Inhibition with Vemurafenib. *Cancer Discovery* **2012**, *2* (3), 227−235.
(52) Mizukami, Y.; Li, J. N.; Zhang, X. B.; Zimmer, M. A.; Iliopoulos, O.; Chung, D. C. Hypoxia-inducible factor-1-independent regulation of vascular endothelial growth factor by hypoxia in colon cancer. *Cancer Res.* **2004**, *64* (5), 1765−1772.
(53) Rahman, K. T.; Islam, M. F.; Banik, R. S.; Honi, U.; Diba, F. S.; Sumi, S. S.; Kabir, S. M. T.; Akhter, M. S. Changes in protein interaction networks between normal and cancer conditions: Total chaos or ordered disorder? *Network Biology* **2013**, *3* (1), 15.
(54) Rosen, L. S.; Jacobs, I. A.; Burkes, R. L. Bevacizumab in Colorectal Cancer: Current Role in Treatment and the Potential of Biosimilars. *Target Oncol* **2017**, *12* (5), 599−610.
(55) Verdaguer, H.; Tabernero, J.; Macarulla, T. Ramucirumab in metastatic colorectal cancer: evidence to date and place in therapy. *Ther. Adv. Med. Oncol.* **2016**, *8* (3), 230−242.
(56) Guren, T. K.; Thomsen, M.; Kure, E. H.; Sorbye, H.; Glimelius, B.; Pfeiffer, P.; Osterlund, P.; Sigurdsson, F.; Lothe, I. M. B.; Dalsgaard, A. M.; et al. Cetuximab in treatment of metastatic colorectal cancer: final survival analyses and extended RAS data from the NORDIC-VII study. *Br. J. Cancer* **2017**, *116* (10), 1271−1278.
(57) Van Cutsem, E.; Kohne, C. H.; Hitre, E.; Zaluski, J.; Chien, C. R. C.; Makhson, A.; D'Haens, G.; Pinter, T.; Lim, R.; Bodoky, G.; et al. Cetuximab and Chemotherapy as Initial Treatment for Metastatic Colorectal Cancer. *N. Engl. J. Med.* **2009**, *360* (14), 1408−1417.
(58) Battaglin, F.; Puccini, A.; Djaballah, S. A.; Lenz, H. J. The impact of panitumumab treatment on survival and quality of life in patients with RAS wild-type metastatic colorectal cancer. *Cancer Manage. Res.* **2019**, *11*, 5911−5924.

https://dx.doi.org/10.1021/acs.jpcb.0c09307
J. Phys. Chem. B 2021, 125, 1716−1726

Supporting Information for Publication

Modular Reorganization of Signaling Networks During the Development of Colon Adenoma and Carcinoma

Klára Schulc[1], Zsolt T. Nagy[1], Sebestyén Kamp[2], János Molnár[2], Daniel V. Veres[2], Peter Csermely[1], Borbála M. Kovács[1]

[1]Department of Molecular Biology, Semmelweis University, Budapest 1428, Hungary

[2]Turbine AI, Budapest, Hungary

Table of Contents

Supporting Texts

Text S1. Checking the robustness of the results by adding noise to the data

Microarray data is well known to be often noisy. To improve the quality we collected data over 100 colon samples from each state (normal; adenoma; carcinoma). We chose to analyze parameters that do not rely and require precise measurements. However, we investigated into the robustness of our results by randomly elevating and decreasing half-half of the abundances by 5%. We found that our most relevant results are robust to this noise (see Figures S10 and S11 Tables S2 and S4).

Text S2. The algorithm for the generation of EntOpt images

Entropy-based visualisation (http://apps.cytoscape.org/apps/entoptlayout) was extensively used in this research to visualise the important changes among the normal, adenoma and carcinoma networks with a link weight sensitive method. The algorithm for the optimal usage of the EntOpt Layout program was described by a member of our research group, Andrea Császár.

i. Each network was first visualised by the built-in Prefuse Force Directed Layout program of Cytoscape.

ii. The maximal run time was set to be 50 000 seconds.

iii. Finally, the program was runned four times, with different settings:

1) Initialize node positions with: Visible coordinates

   Node parameter to optimize for: Position

2) Initialize node positions with: Entopt coordinates

   Node parameter to optimize for: Width

3) Initialize node positions with: Entopt coordinates

   Node parameter to optimize for: Position

4) Initialize node positions with: Entopt coordinates

   Node parameter to optimize for: Width

Text S3. Investigating into the modular overlap changes

The ModuLand plugin is able to detect highly overlapping modules, as well as to calculate the effective number and degree of modules. After comparing these between the three networks, we have found that although there are interesting differences, they are equivalent to the number of modules. Therefore, after normalization to the same number of modules, the differences in the overlap values disappeared. Calculating the exact number of modules is a current challenge in network science, thus these results (Figures S16-19) are interesting, but not reliable enough.

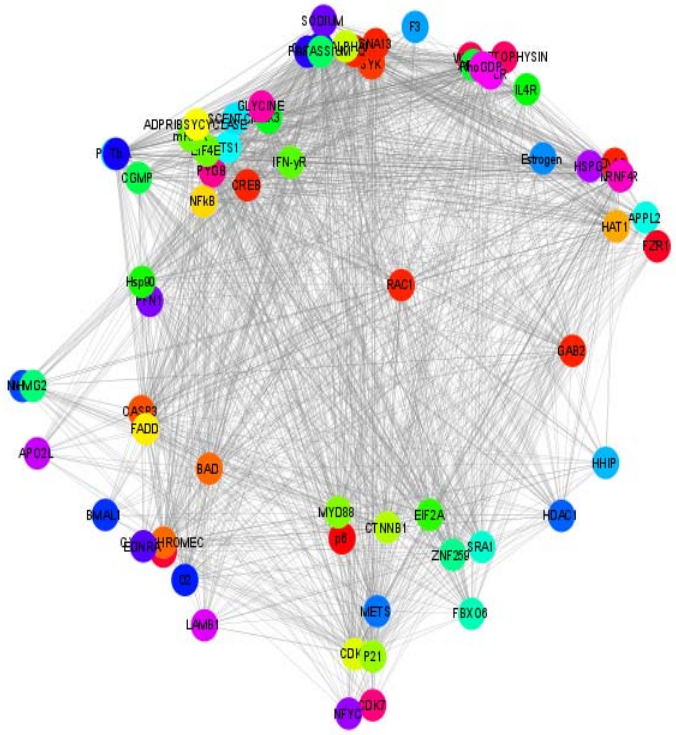
Text S4. The method for the examination of targeted and immunotherapy-related pathways

Targeted and immunotherapy related pathways were analysed with the help of Gene Ontology Consortium, as the proteins belonging to the appropriate GO terms were selected. In case of EGFR and VEGFR inhibitors, ‘epidermal growth factor receptor signaling pathway’ and ‘vascular endothelial growth factor receptor signaling pathway’ terms were used, respectively. In the case of immunotherapy, MSI status and mismatch repair proteins are known to be important predictors of therapy efficacy. In this case, the proteins of the GO term ‘mismatch repair’ were chosen for analysis.

As a next step, after filtering for human proteins, the list of the proteins belonging to the appropriate GO term was downloaded, and the duplications were removed. Then the overlap with our dataset were calculated, and the abundance of the remaining proteins were looked up. Then, the median abundance was calculated for each of the pathways. Their modular affiliation and the relations with the strongest and weakest links in the network were also analysed.

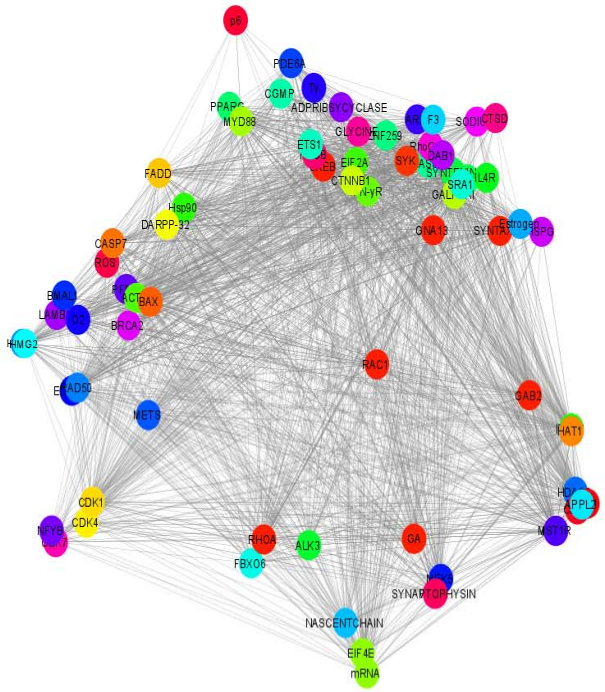
Supporting Figures

Figure S1. The hierarchy of modules in Level 1, calculated with the ModuLand plugin of Cytoscape, with normal link weights



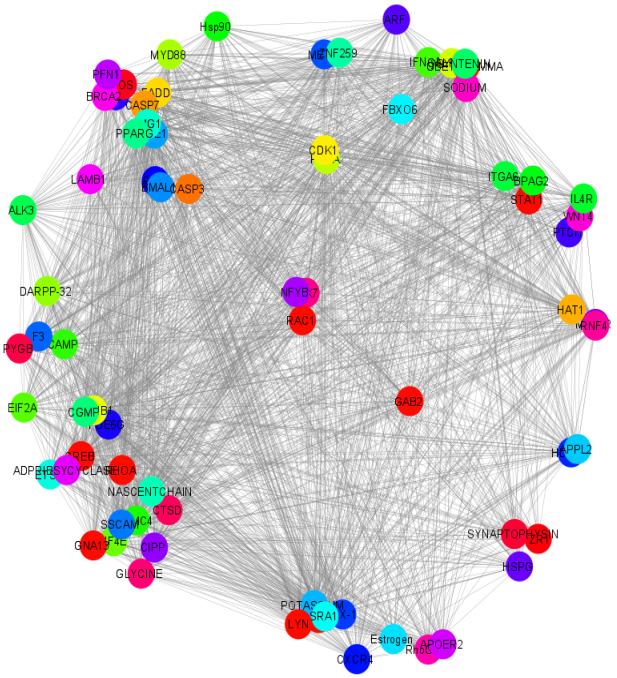
This figure was generated with the ModuLand program. After finding the functional modules in the network, the program generates a hierarchy between them, with the use of the modular link weights. The modules are highlighted with different colors, and the links between them are depicted with grey lines. Normal link weights were used in this calculation.

Figure S2. The hierarchy of modules in Level 1, calculated with the ModuLand plugin of Cytoscape, with adenoma link weights



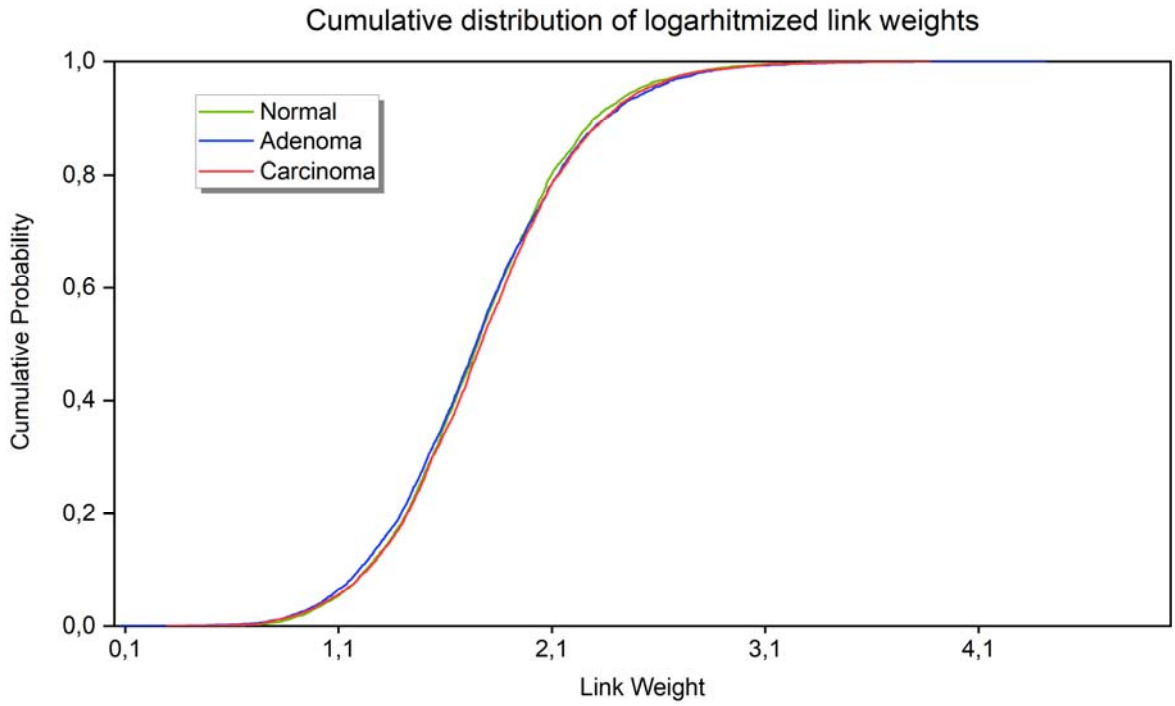
This figure was generated with the ModuLand program. After finding the functional modules in the network, the program generates a hierarchy between them, with the use of the modular link weights. The modules are highlighted with different colors, and the links between them are depicted with grey lines. Adenoma link weights were used in this calculation.

Figure S3. The hierarchy of modules in Level 1, calculated with the ModuLand plugin of Cytoscape, with carcinoma link weights



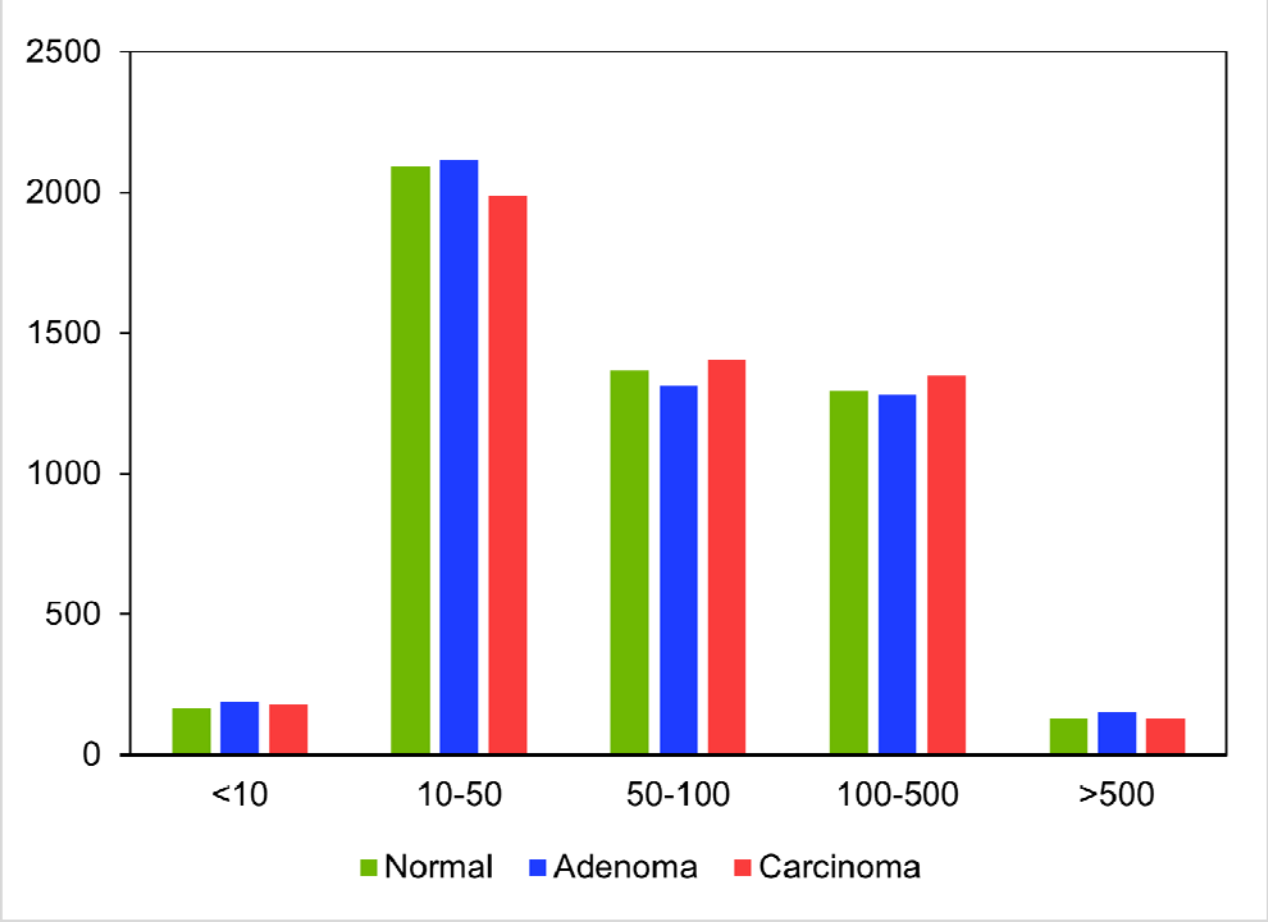
This figure was generated with the ModuLand program. After finding the functional modules in the network, the program generates a hierarchy between them, with the use of the modular link weights. The modules are highlighted with different colors, and the links between them are depicted with grey lines. Carcinoma link weights were used in this calculation.

Figure S4. The distribution of the logarithmic link weights



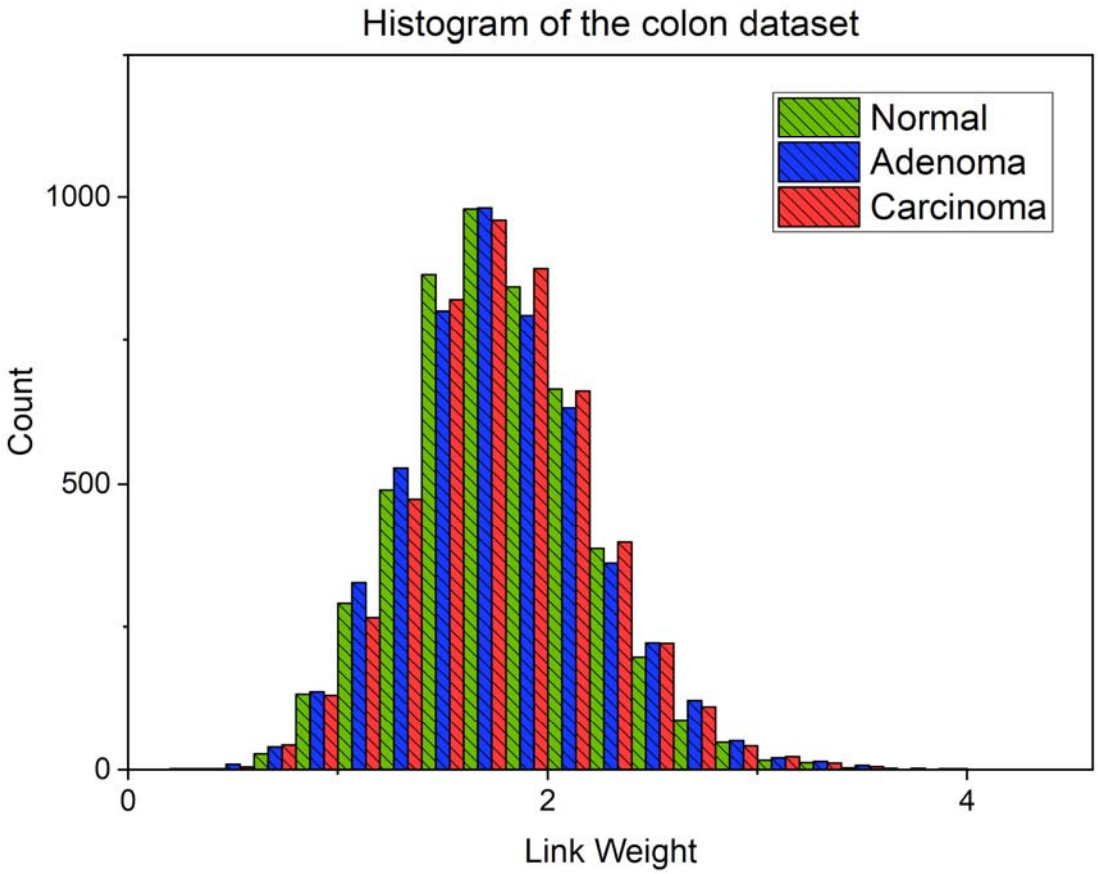
The distribution of the logarithmic link weights demonstrates the large median and standard deviation of the adenoma network, similarly to the box-plot and distribution of the logarithmic and non-logarhitmic values (see the main text).

Figure S5. Number of links in different bins of data



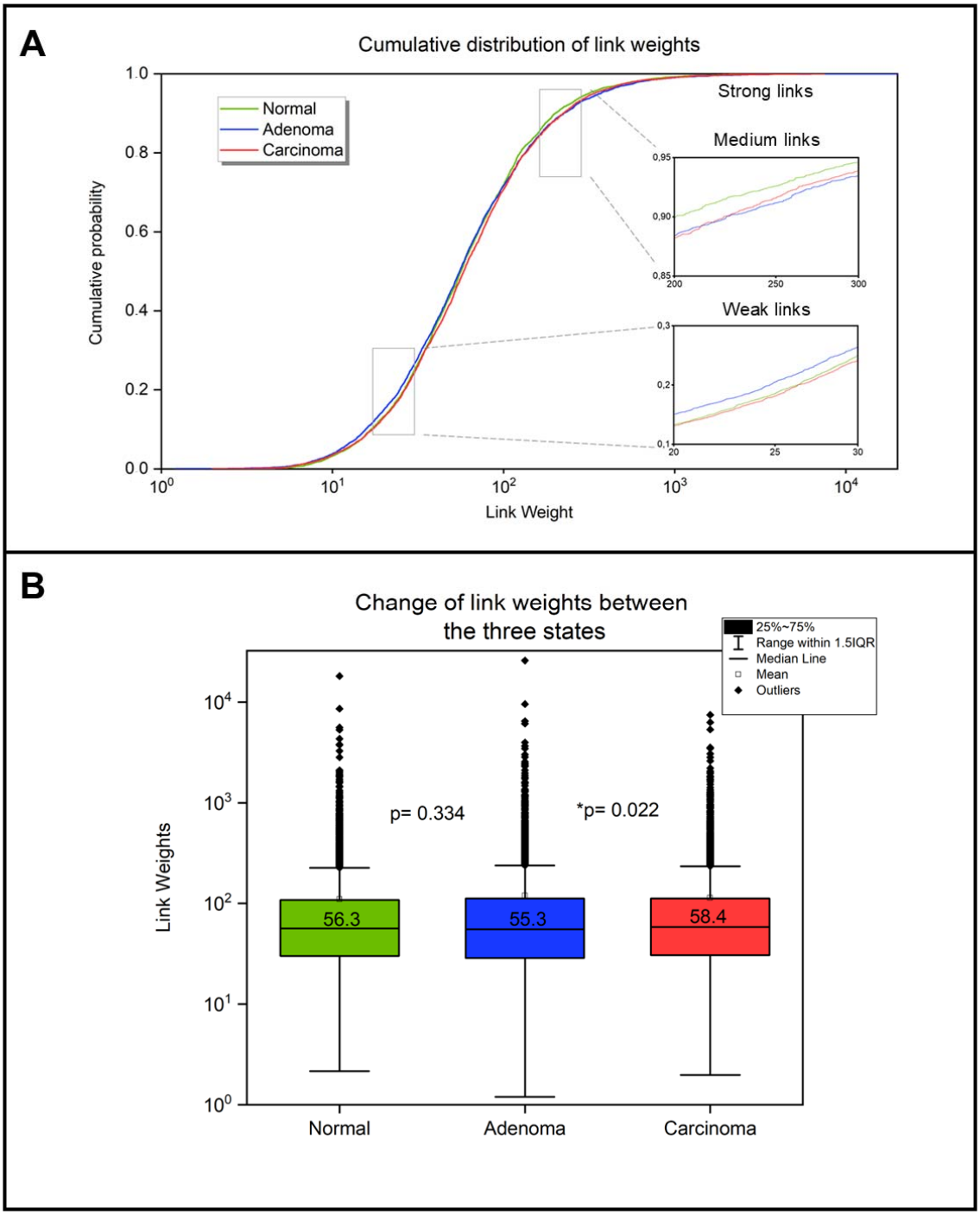
To further demonstrate the differences among the three networks' link weight distribution (see the main text and Figure S6), the link weight data was binned according to the non-logarhitmic link weight cutoff values. It also shows that the adenoma network is the most important among the very small and very large link weights.

Figure S6. Probability density function of the logarithmic link weights



The probability density function of the logarithmic link weight data also demonstrates subtle differences between the distribution of the three networks.

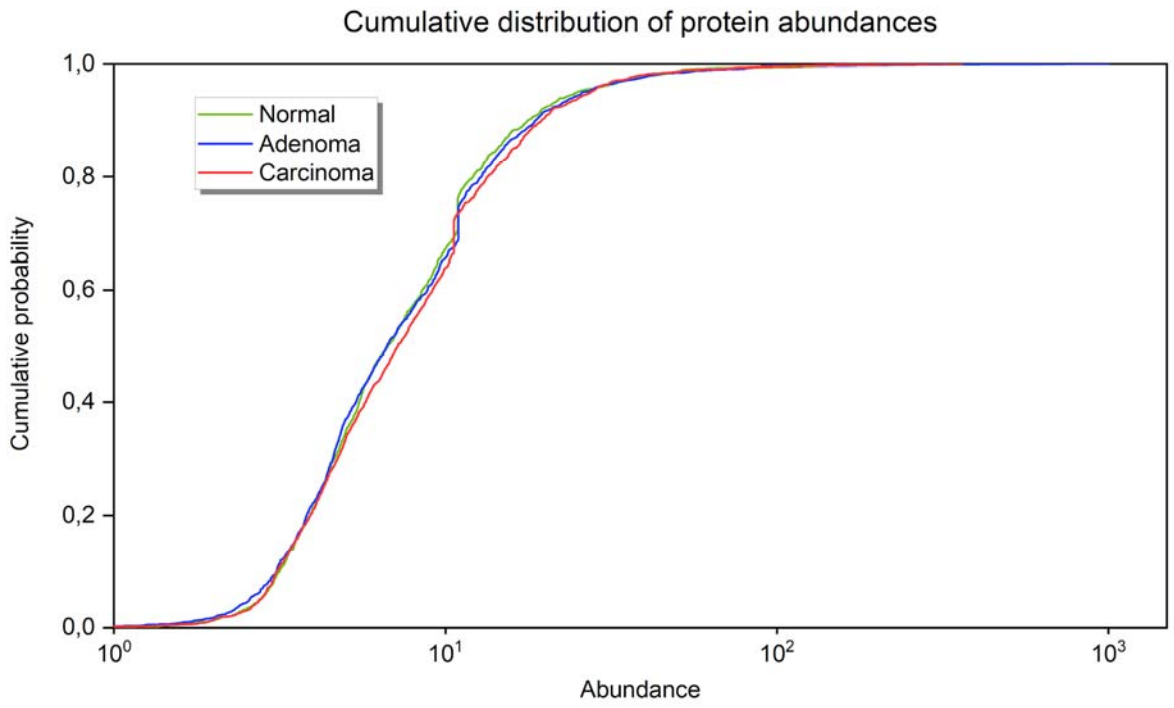
Figure S7. Cumulative distribution and box plot of the non-logarithmic link weights



The cumulative distribution of the non-logarithmic link weights shows the same differences as with the logarithmic values (see the main text). Two areas among the small link weights (defined as link weights under 50) and medium link weights (defined as link weights between 100 and 500) are highlighted.

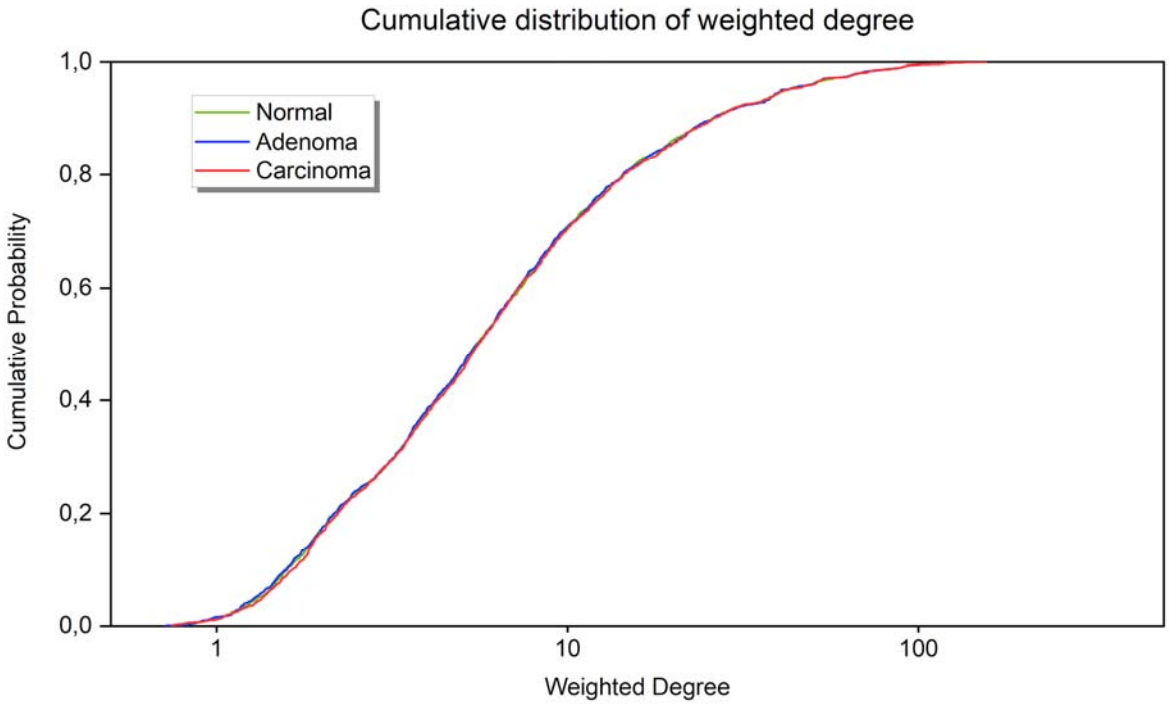
The box-plot of the non-logarithmic link weight demonstrates the large median and standard deviation of the adenoma network, similarly to the box-plot of the logarithmic values (see the main text). The median and p values (paired Wilcoxon) are highlighted.

Figure S8. Cumulative distribution of the abundances



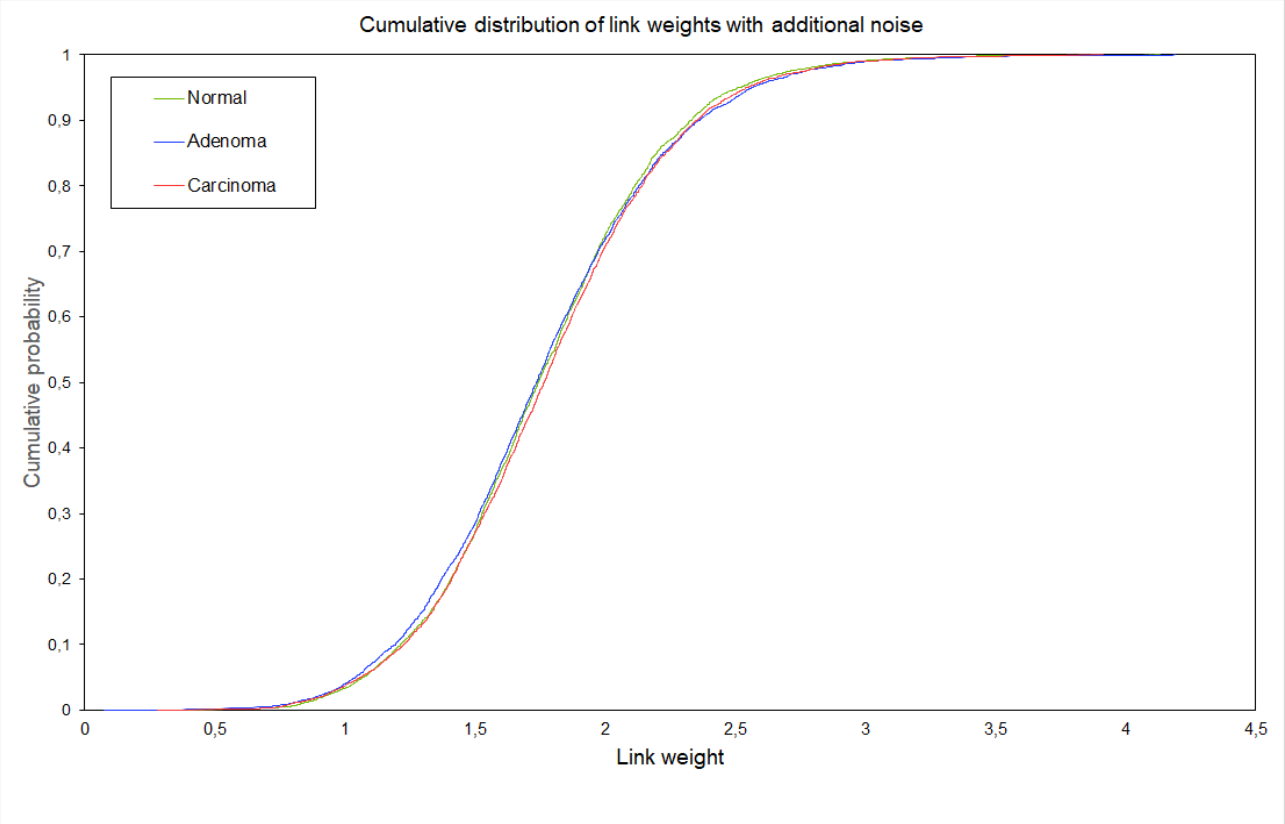
The cumulative distribution of the protein abundances also shows the same main characteristics as the link weight distribution (see the main text). As in the case of a few missing values, the average abundance has been used, a small bulge appears around the average value. This correction did not affect the analysis significantly.

Figure S9. Cumulative distribution of weighted degrees



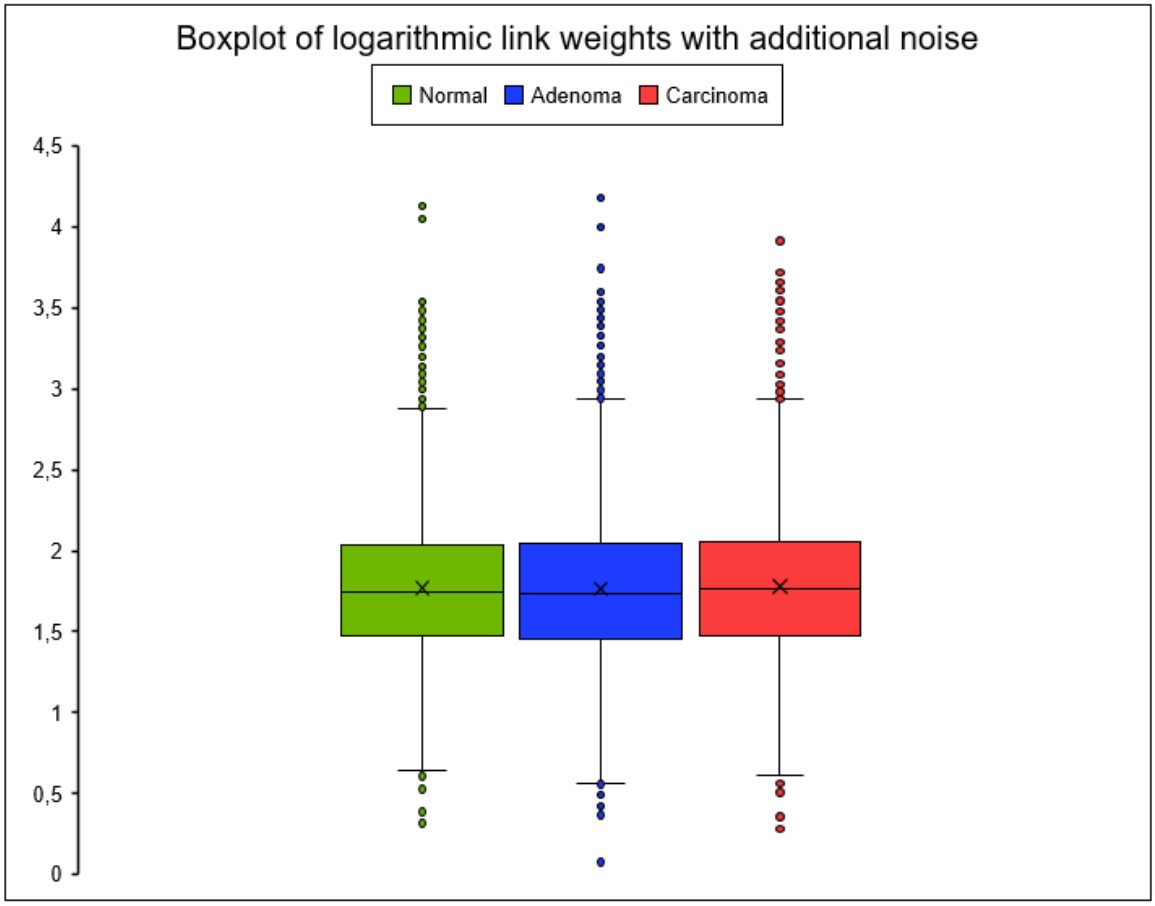
The cumulative distribution of the weighted degrees shows less demonstrable differences than the link weight and abundance distribution (see the main text and Figures S6 and S8). In the area of small and large link weights, the adenoma network has the largest cumulative probability.

Figure S10. Cumulative distribution of link weights with additional 5% noise



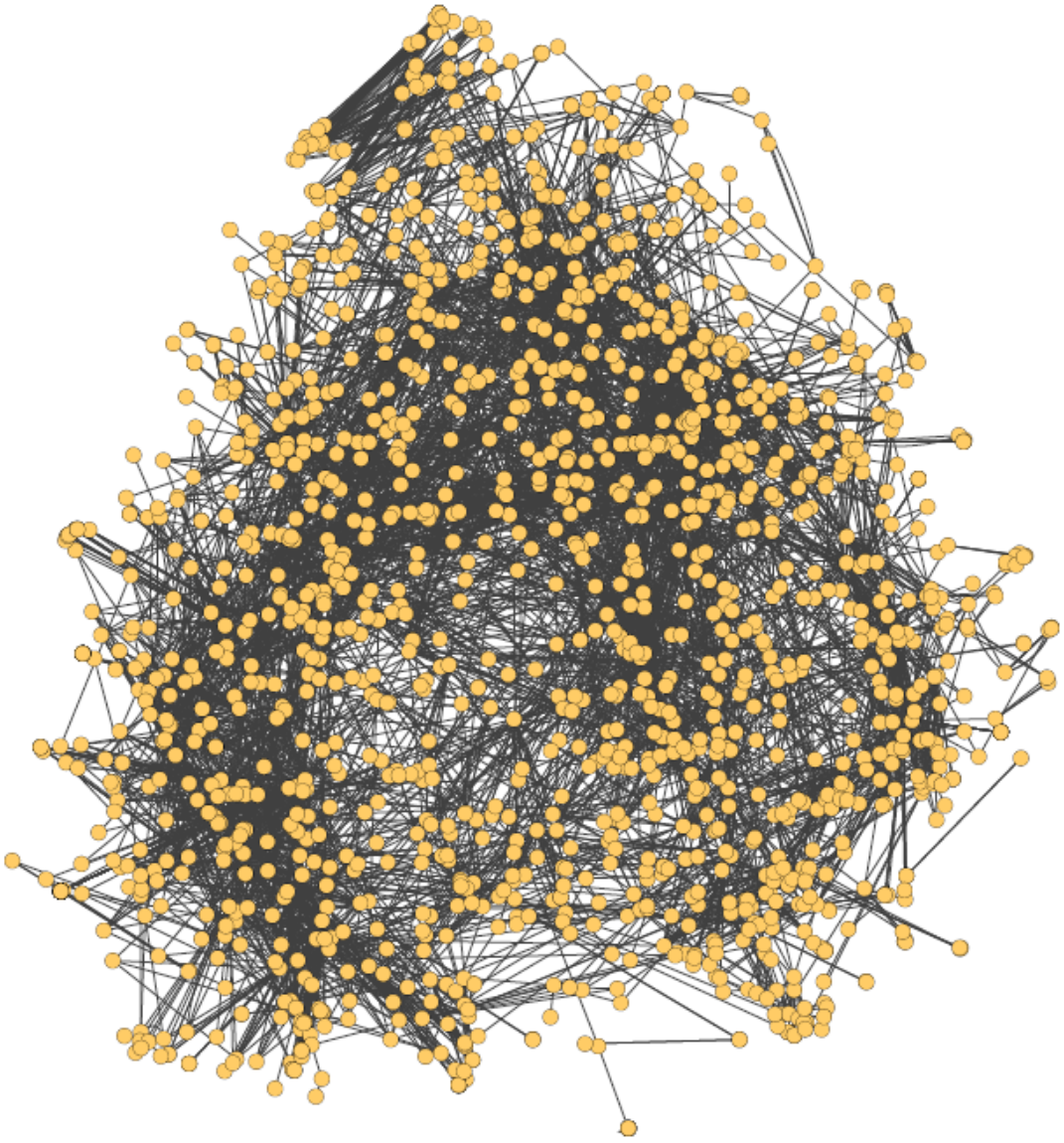
The noisy network was created as described in Text S1. The results are robust to noise, as the adenoma network has the most cumulative probability among the small link weights, and the normal network among the medium link weights (see the main text for context). The results are also significant (paired Wilcoxon-test, $p<0.0001$ for normal-adenoma, adenoma-carcinoma, normal-carcinoma pairs).

Figure S11. Box plot of link weights with additional 5% noise



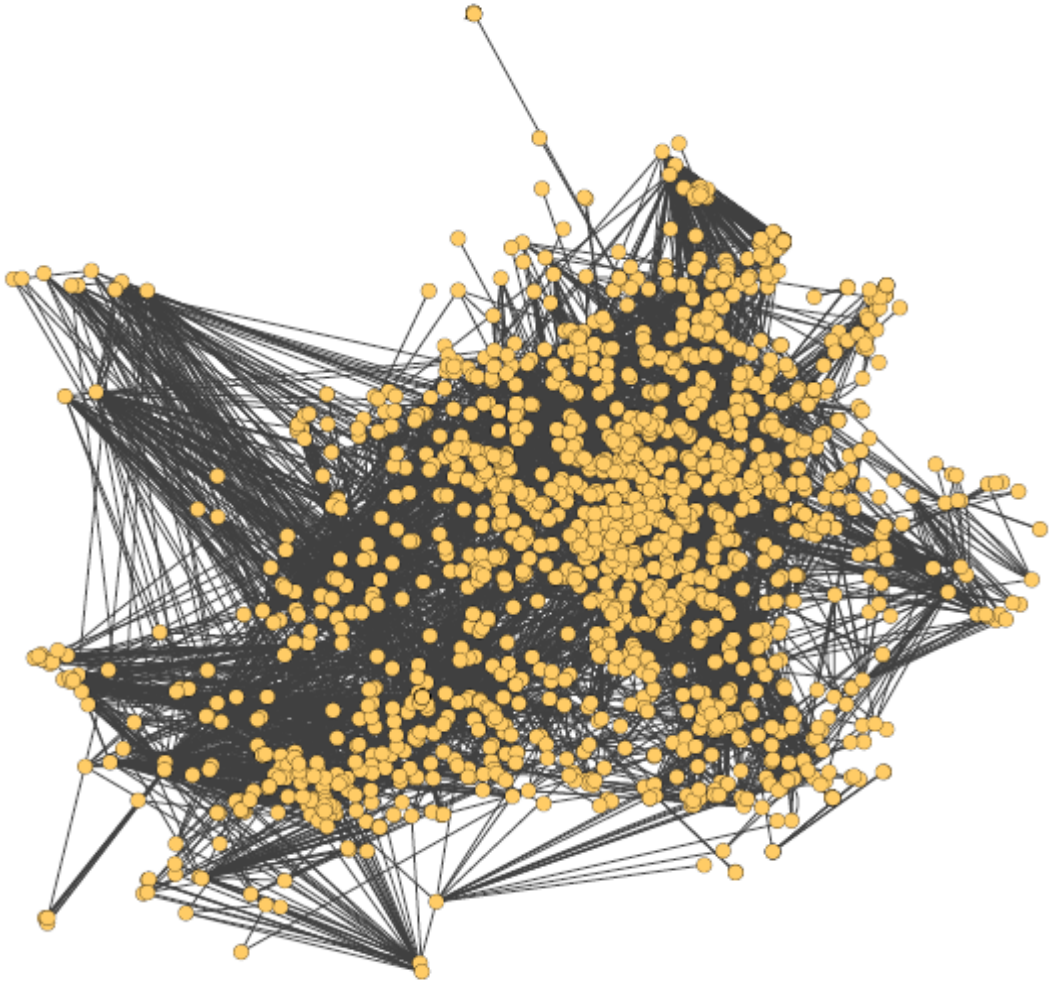
The noisy network was created as described in Text S1. The results are robust to noise, as the adenoma network has the largest, and the carcinoma network has the smallest standard deviation (see the main text for context).

Figure S12. The EntOpt image of the unweighted Human Cancer Signaling Network



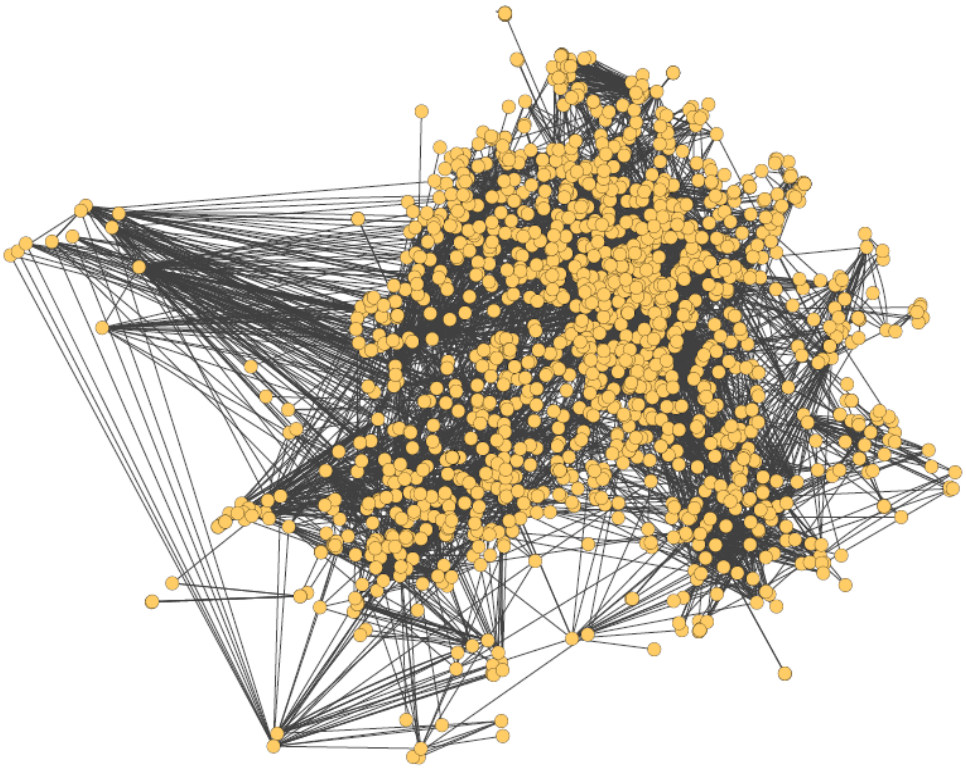
The image was made with the EntOpt Layout program, as described above. The entropy calculations were conducted without the use of link weights. The nodes are highlighted with the color yellow.

Figure S13. The EntOpt image of the network with normal weights



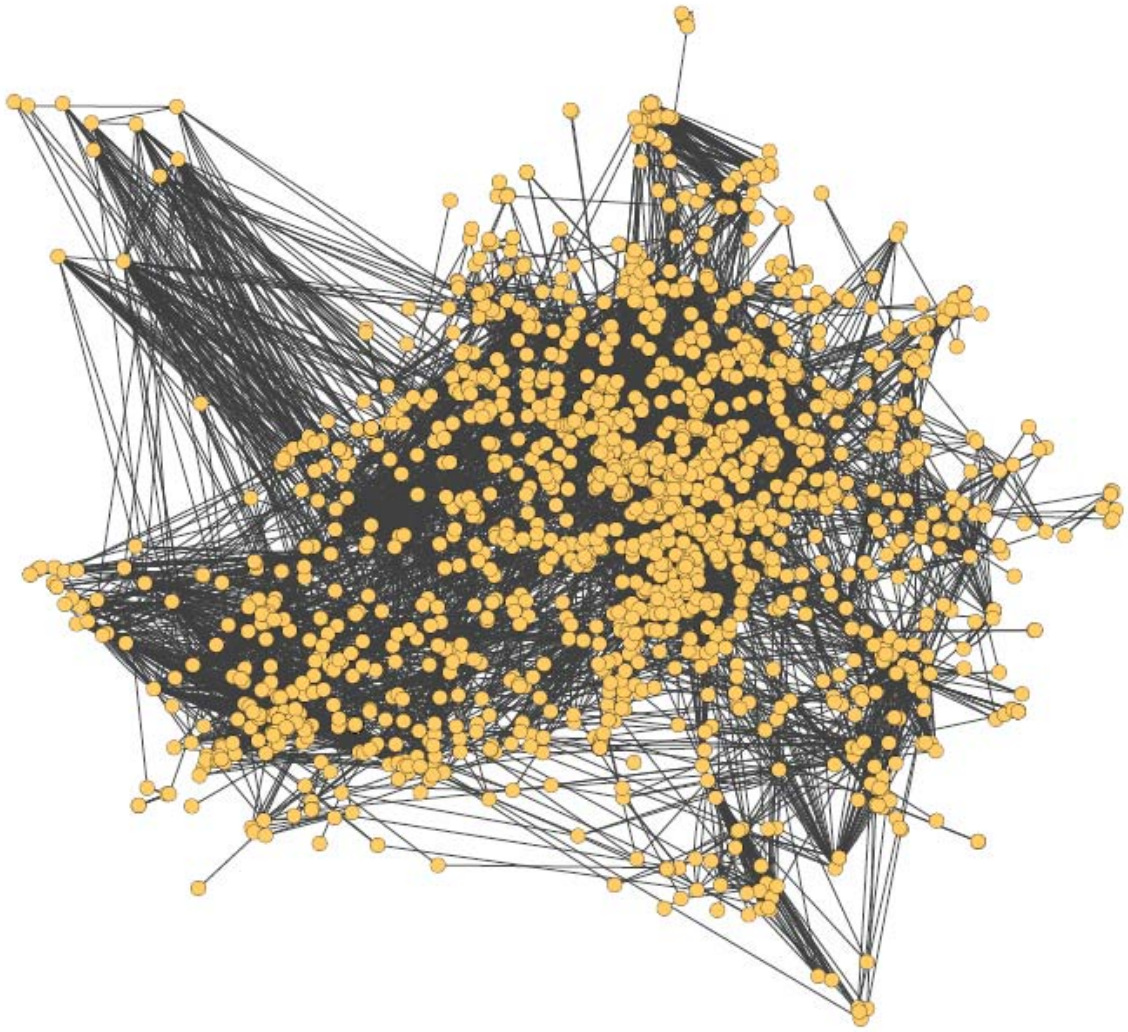
The image was made with the EntOpt Layout program, as described above. The entropy calculations were conducted with the use of the normal colon link weights. The nodes are highlighted with the color yellow.

Figure S14. The EntOpt image of the network with adenoma weights



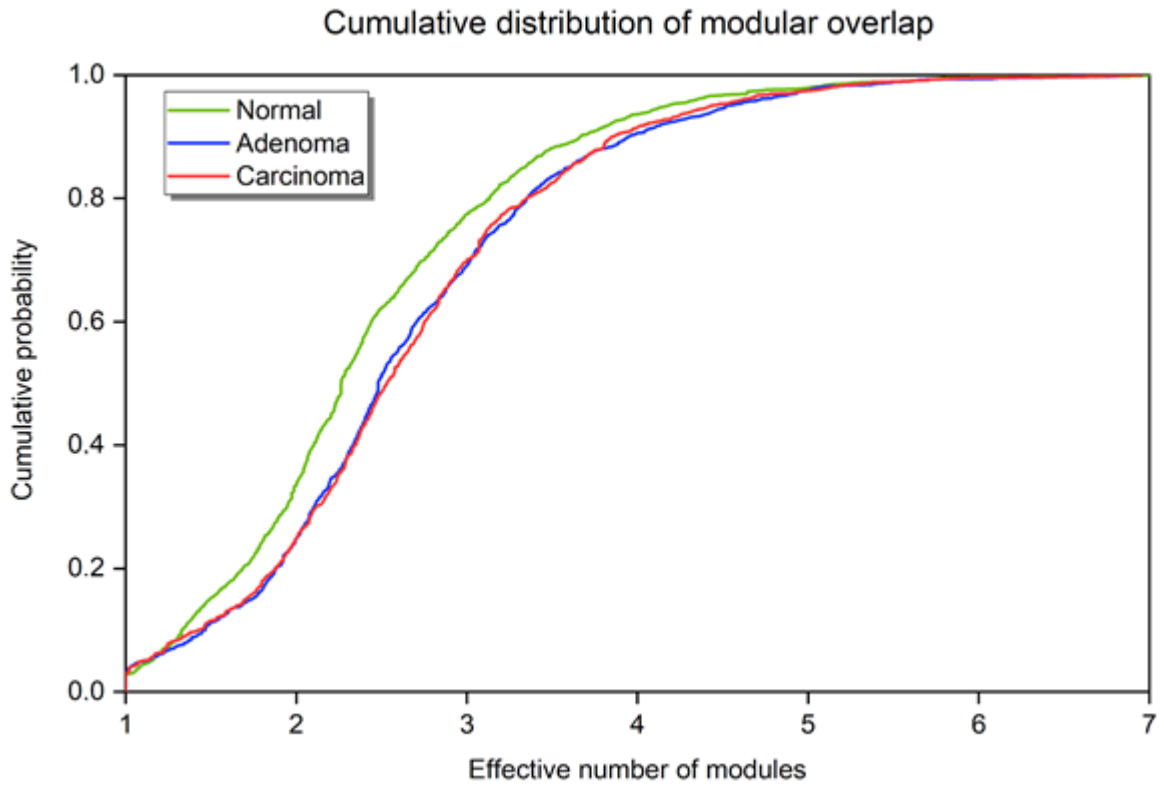
The image was made with the EntOpt Layout program, as described above. The entropy calculations were conducted with the use of the colon adenoma link weights. The nodes are highlighted with the color yellow.

Figure S15. The EntOpt image of the network with carcinoma weights



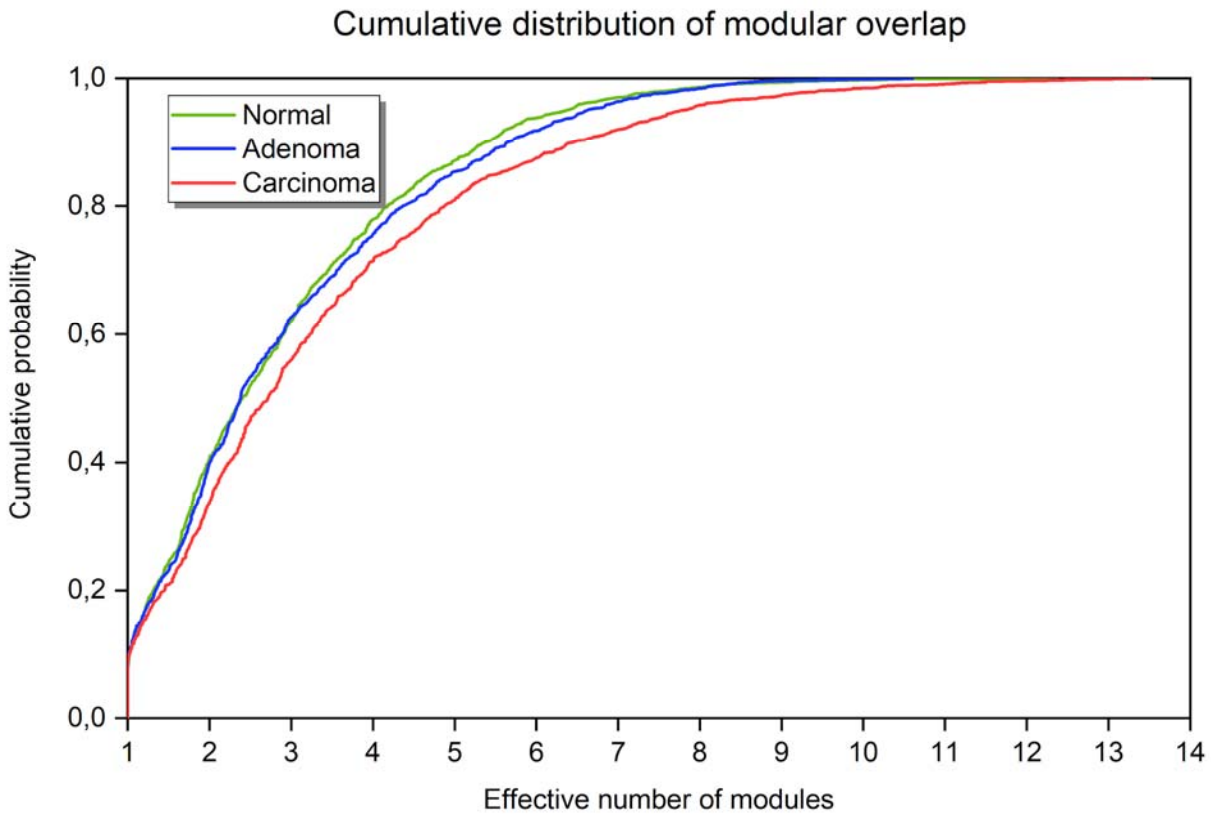
The image was made with the EntOpt Layout program, as described above. The entropy calculations were conducted with the use of the colon carcinoma link weights. The nodes are highlighted with the color yellow.

Figure S16. Change of the ModuLand overlap values based on the different number of modules with logarithmic link weights



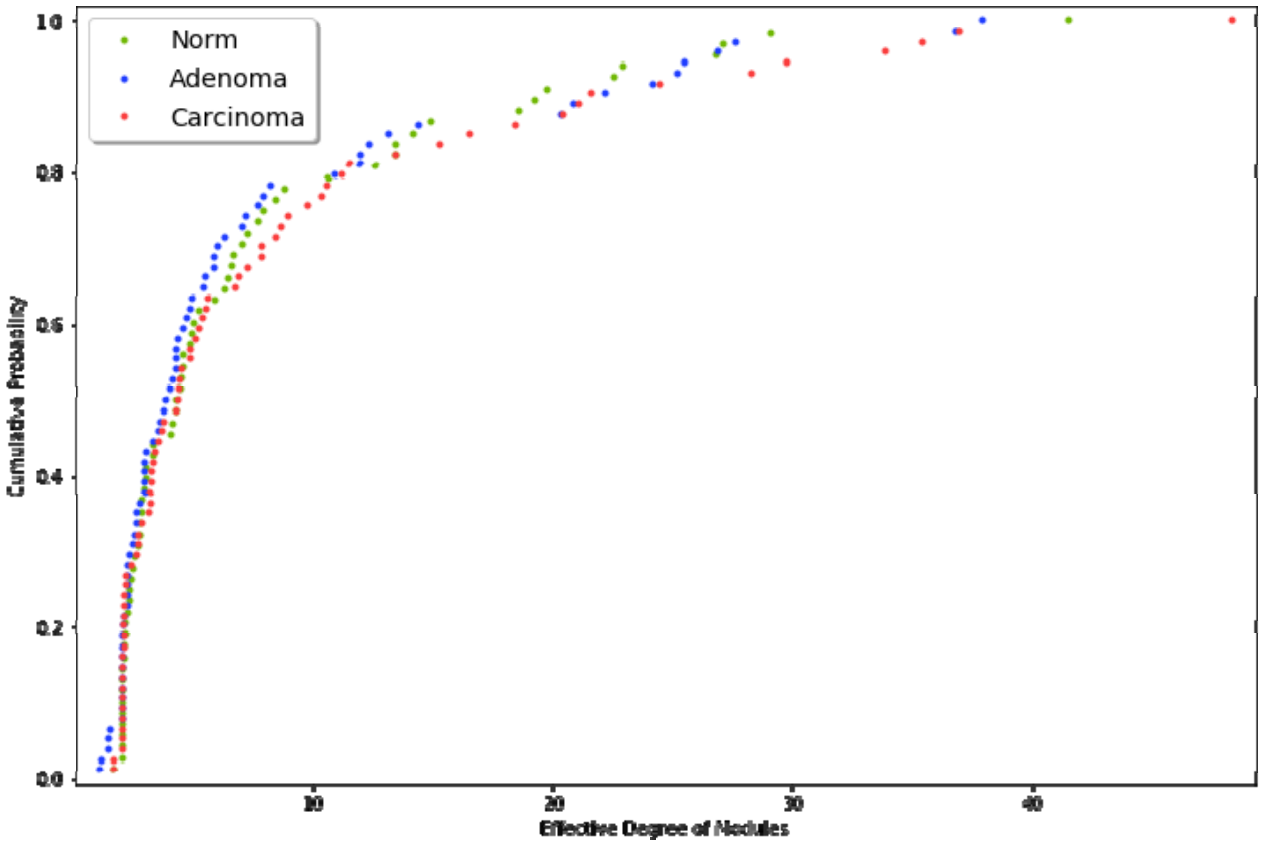
We have investigated into the modular overlap changes, however we found that it strictly relies on the number of modules, which is not a precise entity. Therefore we did not implement this result into the main text.

Figure S17. Change of the ModuLand overlap values based on the different number of modules with non-logarithmic link weights



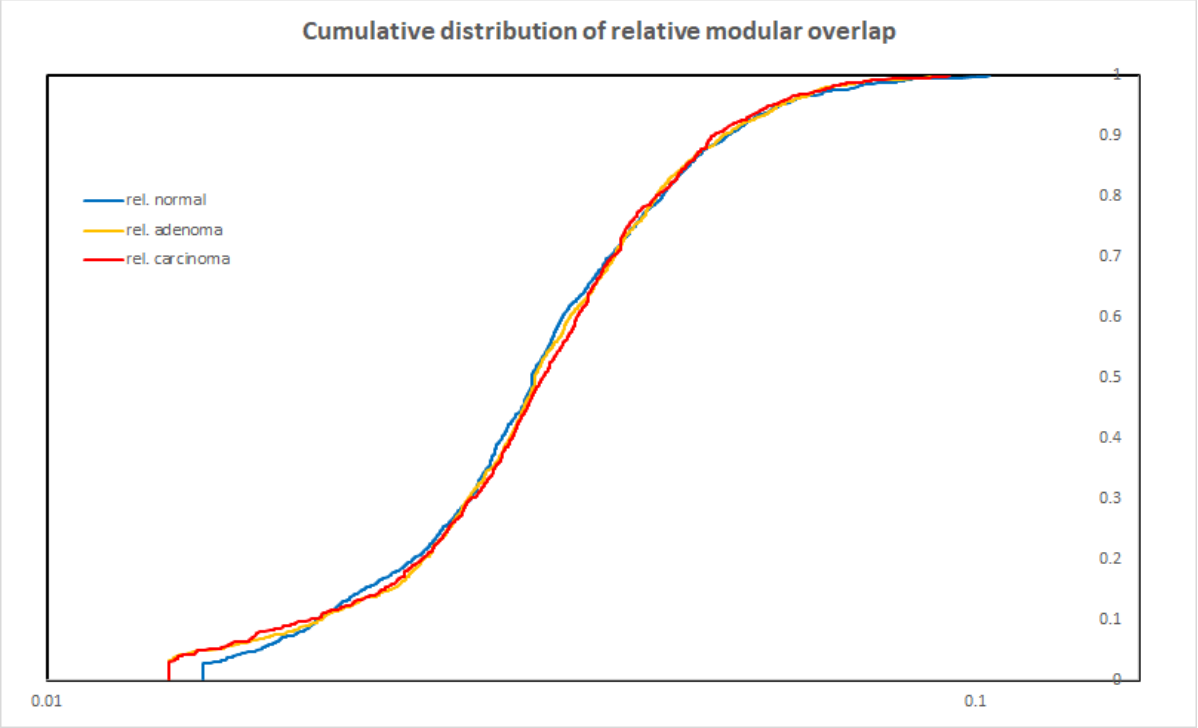
We have investigated into the modular overlap changes, however we found that it strictly relies on the number of modules, which is not a precise entity. Therefore we did not implement this result into the main text.

Figure S18. Cumulative distribution of the effective degree of modules



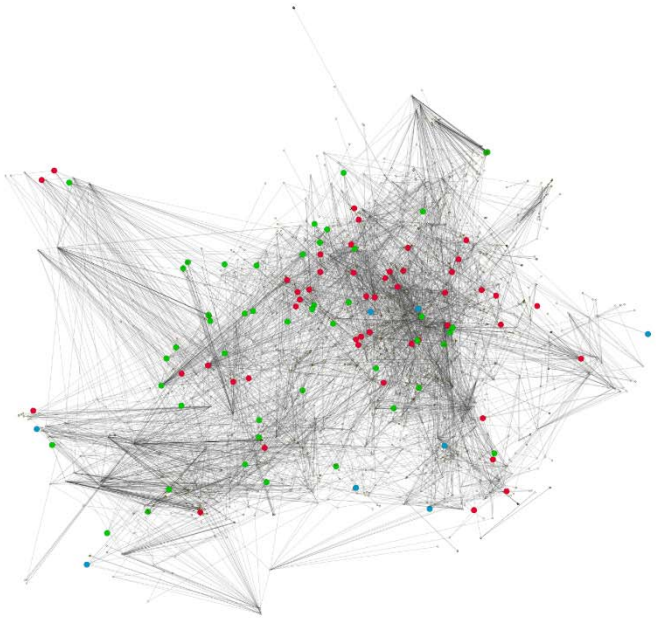
The effective degree of modules were calculated using the ModuLand plugin (see Materials and Methods) from the weighted degree measure of the nodes on the first hierarchical level, each representing a module of the original network. We found that there were no significant changes between the normal, adenoma and carcinoma networks, indicating that the strength of the links between the modules are even. Statistical analysis was performed using Wilcoxon-test. ($p_{N\text{-}A}$ = 0.9723, $p_{N\text{-}C}$ = 0.1101, $p_{A\text{-}C}$= 0.1314)

Figure S19. Cumulative distribution of the normalized modular overlap



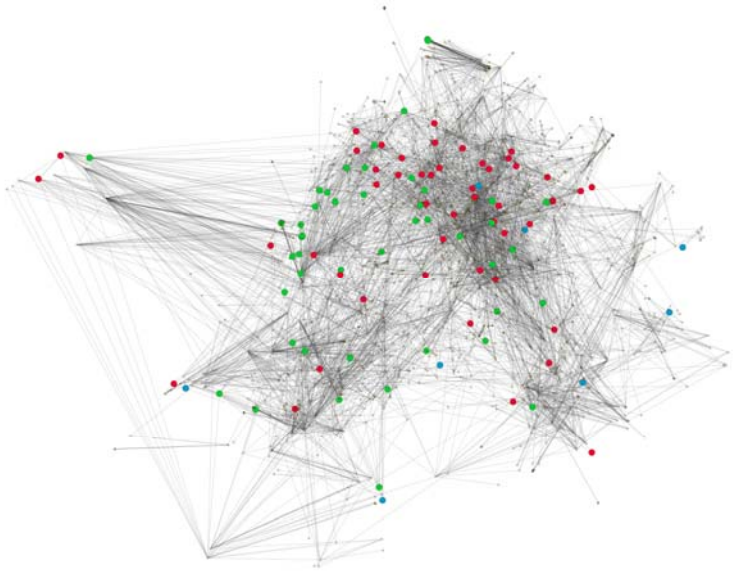
Normalizing the overlap values to the number of modules the difference between normal network versus the adenoma and carcinoma networks almost disappeared, indicating that higher number of modules in the network increases the probability of modular overlap.

Figure S20. EGFR-, VEGFR-signaling and mismatch repair related nodes in the normal network



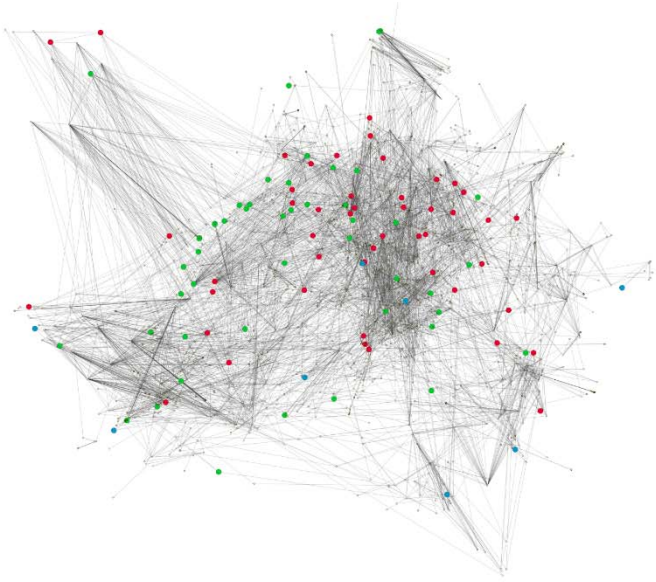
The EntOpt image of the normal network with the coloring of the different pathways. Nodes in the EGFR pathway are green, in the VEGFR pathway are red and in the mismatch repair pathway are blue. As it seems, these nodes do not form different modules, as they are strongly intertwined with the center of the network.

Figure S21. EGFR-, VEGFR-signaling and mismatch repair related nodes in the adenoma network



The EntOpt image of the adenoma network with the coloring of the different pathways. Nodes in the EGFR pathway are green, in the VEGFR pathway are red and in the mismatch repair pathway are blue. As it seems, these nodes do not form different modules, as they are strongly intertwined with the center of the network.

Figure S22. EGFR-, VEGFR-signaling and mismatch repair related nodes in the carcinoma network



The EntOpt image of the carcinoma network with the coloring of the different pathways. Nodes in the EGFR pathway are green, in the VEGFR pathway are red and in the mismatch repair pathway are blue. As it seems, these nodes do not form different modules, as they are strongly intertwined with the center of the network.

Supporting Tables

Table S1. Number of samples in the dataset series

Supporting Table 1. Number of samples in the dataset series

Series	Normal	Adenoma	Carcinoma	Total
GSE20916	44	55	46	145
GSE33113	6	0	90	96
GSE37364	38	29	27	94
GSE4183	8	15	15	38
GSE8671	32	32	0	64
Total	128	131	178	437

In this research, five GEO data series were processed, which contained normal colon, colon adenoma and adenocarcinoma gene expression data. The distribution of the number of samples is shown by the table above.

Table S2. Network diameter calculations with reciprocal and inverted data, and with additional 5% noise

Table S2.

Network diameters in the normal, adenoma and carcinoma network with negative logarhitmic mapping, and 5% added noise

	Undirected[a]		**Directed**[b]		**Mixed graph**[c]	
	Network diameter[d]	**Average path length**[e]	**Network diameter**[d]	**Average path length**[e]	**Network diameter**[d]	**Average path length**[e]
Normal	34.295	*9.885*	37.346	*12.822*	36.435	*11.399*
Adenoma	36.682	*10.573*	40.407	*13.680*	38.876	*12.092*
Carcinoma	29.362	*8.122*	32.779	*10.676*	30.552	*9.314*

Network diameters in the normal, adenoma and carcinoma network with reciprocal and inverted data transformation

	Undirected[a]		**Directed**[b]		**Mixed graph**[c]	
	Reciprocal calculation[f]	**Inverted calculation**[g]	**Reciprocal calculation**[f]	**Inverted calculation**[g]	**Reciprocal calculation**[f]	**Inverted calculation**[g]
Normal	44090	*29260*	23739	*21987*	44447	*29697*
Adenoma	55649	*38724*	32983	*29065*	57834	*38724*
Carcinoma	24523	*20035*	21235	*21235*	31757	*21783*

[a]In the network, every link were assigned as undirected.

[b]The original directivity were preserved in the calculation.

[c]Undirected links were considered as bi-directional links.

[d]Network diameters were calculated with the Dijkstra algorithm.

[e]Average path lengths were calculated with the NetworkX package.

[f]Network diameters were calculated with reciprocal transformation of the data

[g]Network diameters were calculated with inversion of the data

Table S3. Function of the largest network modules nodes

Table S3. Estimated functions of the largest modules				
Name of the module	**Size (number of nodes)**			**Estimated function**
	Normal	***Adenoma***	***Carcinoma***	
RAC1	1000+			signal transduction
FADD	38	33	37	death receptor signaling
GALPHAI	32	27	33	G protein signaling
MYD88	27	28	30	TLR and IL1 signaling (innate immune response)
CDK1	25	35	35	cell cycle continuance
CTNNB1	22	12	16	WNT signaling pathway
GAB2	15	17	9	transcription
EIF4E	15	16	17	translation
BAX		148	159	mitochondrial apoptosis
BAD	128			mitochondrial apoptosis
CASP3	43			apoptosis common pathway
p21	17			cell cycle control
CASP7		13		apoptosis common pathway
CAMP			13	protein kinase A-cAMP signaling
PCNA			8	cell cycle G1-S transition

Table S4. The largest modules of the networks with additional 5% noise

Table S4. Data with additional 5% noise. Modules of the apoptosis and cell cycle merge together

normal		adenoma		carcinoma	
name	size	name	size	name	size
RAC1	990	RAC1	1025	RAC1	1052
BAD	143	p53	75	BAX	169
FADD	37	BAX	47	CDK1	41
CASP3	35	FADD	36	GALPHAI	34
KIT	35	CASP7	32	MYD88	31
GALPHAI	30	MYD88	30	FADD	29
MYD88	28	GALPHAI	27	EIF4E	17
RB	27	CDK1	24	CTNNB1	15
GAB2	20	GAB2	23	ALK3	10
P21	16	PCNA	19	IFN-yR	10
CGMP	13	EIF4E	16	NASCENTCHAIN	10
ALK3	10	HAT1	11		
HAT1	10	ALK3	10		
IFN-yR	10	CTNNB1	10		
NASCENTCHAIN	10	TRAF2	10		

The apoptosis related modules are highlighted with red.
The cell cycle related modules are highlighted with blue.

Tables S5 and S6. The relevant changes in the strongest and weakest 1% of the links

Table S5. The relevant changes in the strongest 1 % of the links

From (Gene name)	To (Gene name)	Link weight (non-logarithmic)	Module source	Module target	Where?	Link weight change	Relation
HAT1	PTP-SL	1048.722	HAT1	HAT1	Normal	weakening in adenoma, slightly in carcinoma	N > C > A
PTP-SL	HAT1	1048.722	HAT1	HAT1	Normal	weakening in adenoma, slightly in carcinoma	N > C > A
ABP1	PICCOLO	2091.075	ABP1	ABP1	Normal	slightly weakening in adenoma and carcinoma	N > A= C
PAR2	GAQ	2002.359	RAC1	RAC1	Normal	slightly weakening in adenoma and carcinoma	N > A= C
ABP1	DYNAMIN	1595.858	ABP1	RAC1	Normal	slightly weakening in adenoma and carcinoma	N > A= C
cIAP2	CASP7	1324.428	CASP3	CASP3	Normal	slightly weakening in adenoma and carcinoma	N > A= C
FAS	VIL2	1277.447	FADD	RAC1	Normal	slightly weakening in adenoma and carcinoma	N > A= C
cIAP1	CASP7	1184.928	CASP3	CASP3	Normal	slightly weakening in adenoma and carcinoma	N > A= C
SHP1	VAV3	1149.772	GAB2	GAB2	Normal	slightly weakening in adenoma and carcinoma	N > A= C
PKAc	DARPP-32	1134.145	BAD	BAD	Normal	slightly weakening in adenoma and carcinoma	N > A= C
APC	CDH1	1102.694	CTNNB1	RAC1	Normal	slightly weakening in adenoma and carcinoma	N > A= C
MYD88	TLR3	1023.805	MYD88	MYD88	Normal	slightly weakening in adenoma and carcinoma	N > A= C
Calpastatin	VIL2	998.885	RAC1	RAC1	Normal	slightly weakening in adenoma and carcinoma	N > A= C
MAPK13	CEBPA	998.260	RAC1	RAC1	Normal	slightly weakening in adenoma and carcinoma	N > A= C
XIAP	CASP7	957.409	CASP3	CASP3	Normal	slightly weakening in adenoma and carcinoma	N > A= C
CASP7	PROKR1	950.368	CASP3	CASP3	Normal	slightly weakening in adenoma and carcinoma	N > A= C
CASP3	CASP7	922.108	CASP3	CASP3	Normal	slightly weakening in adenoma and carcinoma	N > A= C
CASP3	CASP7	922.108	CASP3	CASP3	Normal	slightly weakening in adenoma and carcinoma	N > A= C
CSNK1D	EPS8	881.357	RAC1	RAC1	Normal	slightly weakening in adenoma and carcinoma	N > A= C
RSK	HAT1	867.247	RAC1	HAT1	Normal	slightly weakening in adenoma and carcinoma	N > A= C
IFN-yR	TIP1	863.428	IFN-yR	IFN-yR	Normal	slightly weakening in adenoma and carcinoma	N > A= C
MEK6	MAPK13	853.835	RAC1	RAC1	Normal	slightly weakening in adenoma and carcinoma	N > A= C
CLCA1	ITGB4	5597.775	RAC1	RAC1	Normal	slightly weakening in carcinoma	N = A > C
BETAARRESTIN	PAR2	1840.333	RAC1	RAC1	Normal	slightly weakening in carcinoma	N = A > C
SHC	CEACAM1	1813.733	RAC1	RAC1	Normal	slightly weakening in carcinoma	N = A > C
CDH1	SMAD3	1715.162	RAC1	RAC1	Normal	slightly weakening in carcinoma	N = A > C
PAR2	GBETAGAMMA	1634.350	RAC1	RAC1	Normal	slightly weakening in carcinoma	N = A > C

EPS8	TCF4	1459.934	RAC1	GAB2	Normal	slightly weakening in carcinoma	N = A > C
ERT	EF1A	1103.203	RAC1	RAC1	Normal	slightly weakening in carcinoma	N = A > C
DARPP-32	PPP1CC	1029.393	BAD	BAD	Normal	slightly weakening in carcinoma	N = A > C
CD151	ITGA6	1788.993	RAC1	ITGA6	Adenoma	slightly weakening in normal and carcinoma	A > N = C
DARPP-32	PP1	1584.007	DARPP-32	DARPP-32	Adenoma	slightly weakening in normal and carcinoma	A > N = C
BPAG2	ITGA6	1311.550	RAC1	ITGA6	Adenoma	slightly weakening in normal and carcinoma	A > N = C
PRKCD	ITGA6	1205.886	RAC1	ITGA6	Adenoma	slightly weakening in normal and carcinoma	A > N = C
HES1	ID1	1109.190	RAC1	RAC1	Adenoma	slightly weakening in normal and carcinoma	A > N = C
CASP6	LMNB2	1030.340	CASP7	CASP7	Adenoma	slightly weakening in normal and carcinoma	A > N = C
CCNB1	CDK1	3663.603	CDK1	CDK1	Adenoma	slightly weakening in normal	A = C > N
CCNB1	CDK1	3663.603	CDK1	CDK1	Adenoma	slightly weakening in normal	A = C > N
CDK1	CCNB1	3663.603	CDK1	CDK1	Adenoma	slightly weakening in normal	A = C > N
LAMR1	ITGA6	2558.331	ITGA6	ITGA6	Adenoma	slightly weakening in normal	A = C > N
p53	BIK	2347.750	BAX	BAX	Adenoma	slightly weakening in normal	A = C > N
CCNA2	CDK1	1734.955	CDK1	CDK1	Adenoma	slightly weakening in normal	A = C > N
CCNA2	CDK1	1734.955	CDK1	CDK1	Adenoma	slightly weakening in normal	A = C > N
HMG1	TopoII	1603.390	HMG1	HMG1	Adenoma	slightly weakening in normal	A = C > N
CDK4	PCNA	1603.026	CDK4	CDK4	Adenoma	slightly weakening in normal	A = C > N
PCNA	CDK4	1603.026	CDK4	CDK4	Adenoma	slightly weakening in normal	A = C > N
CD44	EPS8	1320.956	RAC1	RAC1	Adenoma	slightly weakening in normal	A = C > N
BLNK	LYN	1308.223	SYK	SYK	Adenoma	slightly weakening in normal	A = C > N
RSK2	HAT1	1194.485	RAC1	HAT1	Adenoma	slightly weakening in normal	A = C > N
EPS8	RUVBL1	1151.805	RAC1	RAC1	Adenoma	slightly weakening in normal	A = C > N
HAT1	UXT	1065.008	HAT1	HAT1	Adenoma	slightly weakening in normal	A = C > N
UXT	HAT1	1065.008	HAT1	HAT1	Adenoma	slightly weakening in normal	A = C > N
SYNDECAN	SYNTENIN	1063.515	SYNTENIN	SYNTENIN	Adenoma	slightly weakening in normal	A = C > N
LMNB1	LMNB2	1016.402	CASP7	CASP7	Adenoma	slightly weakening in normal	A = C > N
CLCA1	ITGB4	6477.399	RAC1	RAC1	Adenoma	slightly weakening in carcinoma	A = N > C
CDH1	SMAD3	1287.473	RAC1	RAC1	Adenoma	slightly weakening in carcinoma	A = N > C
DARPP-32	PPP1CC	1043.032	DARPP-32	BAX	Adenoma	slightly weakening in carcinoma	A = N > C
ERT	EF1A	1005.412	RAC1	RAC1	Adenoma	slightly weakening in carcinoma	A = N > C
EPS8	TCF4	999.437	RAC1	GAB2	Adenoma	slightly weakening in carcinoma	A = N > C
PAR2	GBETAGAMMA	973.656	RAC1	RAC1	Adenoma	slightly weakening in carcinoma	A = N > C
CCNA2	CDK4	1019.792	CDK1	PCNA	Carcinoma	weakening in normal, slightly in adenoma	C > A > N
HIF1A	Noxa	1107.099	BAX	BAX	Carcinoma	weakening in normal, slightly in adenoma	C > A > N
CDK1	PLK1	1343.745	CDK1	CDK1	Carcinoma	weakening in normal, slightly in adenoma	C > A > N
IL1A	IL8	1347.597	FADD	RAC1	Carcinoma	weakening in normal and adenoma	C >> A = N
RELA	IL8	2210.242	RAC1	RAC1	Carcinoma	weakening in normal, slightly in adenoma	C > A > N
HES1	PLSCR1	971.908	RAC1	RAC1	Carcinoma	slightly weakening in normal and adenoma	C > A = N
STAT1	PLSCR1	1172.367	RAC1	RAC1	Carcinoma	slightly weakening in normal and adenoma	C > A = N
Importin	RAN	1226.940	RAC1	RAC1	Carcinoma	slightly weakening in normal and adenoma	C > A = N

Importin	RAN	1226.940	RAC1	RAC1	Carcinoma	slightly weakening in normal and adenoma	C > A = N
PLSCR1	PTPN12	1229.324	RAC1	RAC1	Carcinoma	slightly weakening in normal and adenoma	C > A = N
PTPN12	PLSCR1	1229.324	RAC1	RAC1	Carcinoma	slightly weakening in normal and adenoma	C > A = N
CCND1	PCNA	1283.714	PCNA	PCNA	Carcinoma	slightly weakening in normal and adenoma	C > A = N
HAT1	UXT	949.671	HAT1	HAT1	Carcinoma	slightly weakening in normal	C = A > N
UXT	HAT1	949.671	HAT1	HAT1	Carcinoma	slightly weakening in normal	C = A > N
BLNK	LYN	967.056	LYN	RAC1	Carcinoma	slightly weakening in normal	C = A > N
p53	BIK	1022.720	BAX	BAX	Carcinoma	slightly weakening in normal	C = A > N
SYNDECAN	SYNTENIN	1060.180	SYNTENIN	SYNTENIN	Carcinoma	slightly weakening in normal	C = A > N
RSK2	HAT1	1068.930	RAC1	HAT1	Carcinoma	slightly weakening in normal	C = A > N
LMNB1	LMNB2	1146.680	BAX	BAX	Carcinoma	slightly weakening in normal	C = A > N
CD44	EPS8	1196.907	RAC1	RAC1	Carcinoma	slightly weakening in normal	C = A > N
EPS8	RUVBL1	1268.265	RAC1	RAC1	Carcinoma	slightly weakening in normal	C = A > N
LAMR1	ITGA6	1421.976	ITGA6	ITGA6	Carcinoma	slightly weakening in normal	C = A > N
HMG1	TopoII	1803.851	HMG1	HMG1	Carcinoma	slightly weakening in normal	C = A > N
CDK4	PCNA	1842.590	PCNA	PCNA	Carcinoma	slightly weakening in normal	C = A > N
PCNA	CDK4	1842.590	PCNA	PCNA	Carcinoma	slightly weakening in normal	C = A > N
CCNA2	CDK1	2834.158	CDK1	CDK1	Carcinoma	slightly weakening in normal	C = A > N
CCNA2	CDK1	2834.158	CDK1	CDK1	Carcinoma	slightly weakening in normal	C = A > N
CCNB1	CDK1	5345.275	CDK1	CDK1	Carcinoma	slightly weakening in normal	C = A > N
CCNB1	CDK1	5345.275	CDK1	CDK1	Carcinoma	slightly weakening in normal	C = A > N
CDK1	CCNB1	5345.275	CDK1	CDK1	Carcinoma	slightly weakening in normal	C = A > N

Supporting Table 6. The relevant changes in the weakest 1% of the links

From (Gene name)	To (Gene name)	Link weight (non-logarithmic)	Module source	Module target	Where?	Link weight change	Relation
AR	WNT2	5.999	RAC1	RAC1	Normal	strengthening in carcinoma, slightly in adenoma	C > A > N
EGR2	BNIP3	6.406	BAD	BAD	Normal	strengthening in carcinoma, slightly in adenoma	C > A > N
RELN	APOER2	6.473	VLDLR	VLDLR	Normal	strengthening in carcinoma, slightly in adenoma	C > A > N
AGTR2	AGT	6.868	RAC1	RAC1	Normal	strengthening in adenoma and carcinoma	C = A >> N
SYNAPTOTAGMIN	LTYPECA	4.339	RAC1	RAC1	Normal	slightly strengthening in carcinoma	C > A = N
CAMK2	IL1A	5.838	RAC1	FADD	Normal	slightly strengthening in adenoma and carcinoma	C = A > N
BMP-7	BMPR1	5.982	ALK3	ALK3	Normal	slightly strengthening in adenoma and carcinoma	C = A > N
WIF1	WNT	6.063	RAC1	RAC1	Normal	slightly strengthening in adenoma and carcinoma	C = A > N
AR	RAC3	6.781	RAC1	RAC1	Normal	slightly strengthening in adenoma and carcinoma	C = A > N
WNT	AR	6.805	RAC1	RAC1	Normal	slightly strengthening in adenoma and carcinoma	C = A > N
AMPHIPHYSIN	DYNAMIN	6.835	RAC1	RAC1	Normal	slightly strengthening in adenoma and carcinoma	C = A > N
cAMPGEFII	RIM	6.873	RAC1	RAC1	Normal	slightly strengthening in adenoma and carcinoma	C = A > N
cAMPGEFII	RIM	6.873	RAC1	RAC1	Normal	slightly strengthening in adenoma and carcinoma	C = A > N
AR	SLC25A4	6.889	RAC1	RAC1	Normal	slightly strengthening in adenoma and carcinoma	C = A > N
ANKYRIN	NRCAM	6.894	RAC1	RAC1	Normal	slightly strengthening in adenoma and carcinoma	C = A > N
NMDAR	RACK	6.909	RAC1	RAC1	Normal	slightly strengthening in adenoma and carcinoma	C = A > N
JNK3	RNPK	6.923	RAC1	RAC1	Normal	slightly strengthening in adenoma and carcinoma	C = A > N
GALPHAZ	AC5	6.955	RAC1	RAC1	Normal	slightly strengthening in adenoma and carcinoma	C = A > N
BMP-10	BMPR1	5.445	ALK3	ALK3	Normal	slightly strengthening in adenoma	A > C = N
AR	SRY	6.087	RAC1	RAC1	Normal	slightly strengthening in adenoma	A > C = N
PKA	TAU	6.441	RAC1	RAC1	Normal	slightly strengthening in adenoma	A > C = N
AMPAR	HOMER	6.756	RAC1	RAC1	Normal	slightly strengthening in adenoma	A > C = N
SMAD4	AR	6.815	RAC1	RAC1	Normal	slightly strenghtening in carcinoma	C > A = N
RELN	VLDLR	6.153	DAB1	DAB1	Adenoma	strengthening in normal, slightly in carcinoma	N > C > A
SMAD4	FORKHEAD	5.788	RAC1	RAC1	Adenoma	strengthening in carcinoma, slightly in normal	C > N > A
CryAB	CASP3	3.874	BAX	BAX	Adenoma	slightly strengthening in normal and carcinoma	C = N > A
AR	FORKHEAD	4.418	RAC1	RAC1	Adenoma	slightly strengthening in normal and carcinoma	C = N > A
ABI2	WAVE3	4.562	RAC1	RAC1	Adenoma	slightly strengthening in normal and carcinoma	C = N > A
MAP1B	TUBULIN	4.851	RAC1	RAC1	Adenoma	slightly strengthening in normal and carcinoma	C = N > A
NIK	WAVE1	4.926	RAC1	RAC1	Adenoma	slightly strengthening in normal and carcinoma	C = N > A
PCAF	AR	5.327	RAC1	RAC1	Adenoma	slightly strengthening in normal and carcinoma	C = N > A
JIP	TIAM1	5.330	RAC1	RAC1	Adenoma	slightly strengthening in normal and carcinoma	C = N > A
NMDAR	YOTIAO	5.336	RAC1	RAC1	Adenoma	slightly strengthening in normal and carcinoma	C = N > A
APAF1	WAVE3	5.567	BAX	RAC1	Adenoma	slightly strengthening in normal and carcinoma	C = N > A

WAVE3	APAF1	5.567	RAC1	BAX	Adenoma	slightly strengthening in normal and carcinoma	C = N > A
FYN	NMDAR	5.901	RAC1	RAC1	Adenoma	slightly strengthening in normal and carcinoma	C = N > A
SMAD4	ZFHX1B	5.920	RAC1	RAC1	Adenoma	slightly strengthening in normal and carcinoma	C = N > A
RACK	FYN	6.016	RAC1	RAC1	Adenoma	slightly strengthening in normal and carcinoma	C = N > A
FYN	AMPAR	6.069	RAC1	RAC1	Adenoma	slightly strengthening in normal and carcinoma	C = N > A
AR	GP130	5.175	RAC1	RAC1	Adenoma	slightly strengthening in normal	N > C = A
TRKB	FYN	5.650	RAC1	RAC1	Adenoma	slightly strengthening in normal	N > C = A
PKCA	SNAP25	5.665	RAC1	SYNTAXIN	Adenoma	slightly strengthening in normal	N > C = A
CAMK2A	TIAM1	5.755	RAC1	RAC1	Adenoma	slightly strengthening in normal	N > C = A
ANKYRIN	FASCIN	5.849	RAC1	RAC1	Adenoma	slightly strengthening in normal	N > C = A
SMAD4	AR	5.004	RAC1	RAC1	Adenoma	slightly strengthening in carcinoma	N > C = A
SYNAPTOTAGMIN	LTYPECA	5.826	RAC1	RAC1	Adenoma	slightly strengthening in carcinoma	N > C = A
NTF5	TRKB	5.180	RAC1	RAC1	Carcinoma	slightly strengthening in normal and adenoma	N = A > C
NEUROLIGIN	PSD95	5.181	RAC1	RAC1	Carcinoma	slightly strengthening in normal and adenoma	N = A > C
M2R	GALPHAO	5.451	GALPHAI	GALPHAI	Carcinoma	slightly strengthening in normal and adenoma	N = A > C
CB1R	GALPHAO	5.739	GALPHAI	GALPHAI	Carcinoma	slightly strengthening in normal and adenoma	N = A > C
M4R	EEF1A2	5.775	GALPHAI	GALPHAI	Carcinoma	slightly strengthening in normal and adenoma	N = A > C
AR	CDK9	5.803	RAC1	RAC1	Carcinoma	slightly strengthening in normal and adenoma	N = A > C
NOS1	PSD95	5.852	RAC1	RAC1	Carcinoma	slightly strengthening in normal and adenoma	N = A > C
SSTR2	GALPHAO	6.027	GALPHAI	GALPHAI	Carcinoma	slightly strengthening in normal and adenoma	N = A > C
MOPR	GALPHAO	6.104	GALPHAI	GALPHAI	Carcinoma	slightly strengthening in normal and adenoma	N = A > C
KOPR	GALPHAO	6.125	GALPHAI	GALPHAI	Carcinoma	slightly strengthening in normal and adenoma	N = A > C
BDNF	TRKB	6.156	RAC1	RAC1	Carcinoma	slightly strengthening in normal and adenoma	N = A > C
NOS1	PSD93	6.182	RAC1	RAC1	Carcinoma	slightly strengthening in normal and adenoma	N = A > C
KV12	PSD95	6.208	RAC1	RAC1	Carcinoma	slightly strengthening in normal and adenoma	N = A > C
CHAPSYN110	PSD95	6.309	RAC1	RAC1	Carcinoma	slightly strengthening in normal and adenoma	N = A > C
ANKYRIN	FASCIN	5.103	RAC1	RAC1	Carcinoma	slightly strengthening in normal	N > A = C
PKCA	SNAP25	6.092	RAC1	RAC1	Carcinoma	slightly strengthening in normal	N > A = C
CAMK2A	TIAM1	6.229	RAC1	RAC1	Carcinoma	slightly strengthening in normal	N > A = C
TRKB	FYN	6.307	RAC1	RAC1	Carcinoma	slightly strengthening in normal	N > A = C
BMP-10	BMPR1	5.084	ALK3	ALK3	Carcinoma	slightly strengthening in adenoma	A > C = N
PKA	TAU	5.579	RAC1	RAC1	Carcinoma	slightly strengthening in adenoma	A > C = N
AMPAR	HOMER	5.656	RAC1	RAC1	Carcinoma	slightly strengthening in adenoma	A > C = N
AR	SRY	5.837	RAC1	RAC1	Carcinoma	slightly strengthening in adenoma	A > C = N

Table S7. The relevant changes in the strongest and weakest 1% of the links with additional 5% noise

Table S7. Top 1% of linkweights with additional 5% of noise

Normal				
From (Gene name)	**To (Gene name)**	**Noisy_linkweight**	**Module_source**	**Module_target**
CASP6	CASP7	3.400	**CASP3**	**CASP3**
CASP6	LMNB1	3.316	**CASP3**	**CASP3**
cIAP2	CASP7	2.966	**CASP3**	**CASP3**
PKAc	DARPP-32	3.207	**BAD**	**BAD**
cIAP1	CASP7	2.920	**CASP3**	**CASP3**
DARPP-32	PPP1CC	3.163	**BAD**	**BAD**
TFF1	APAF1	2.901	**RAC1**	**BAD**
CASP3	CASP7	3.113	**CASP3**	**CASP3**
XIAP	CASP7	2.832	**CASP3**	**CASP3**
CASP7	PROKR1	2.829	**CASP3**	**CASP3**
CASP3	CASP7	2.816	**CASP3**	**CASP3**
Adenoma				
From (Gene name)	**To (Gene name)**	**Noisy_linkweight**	**Module_source**	**Module_target**
CCNB1	CDK1	3.742	**CDK1**	**CDK1**
CDK1	CCNB1	3.386	**CDK1**	**CDK1**
CCNB1	CDK1	3.386	**CDK1**	**CDK1**
CASP6	CASP7	3.559	**CASP7**	**CASP7**
p53	BIK	3.539	**p53**	**p53**
CASP6	LMNB1	3.491	**CASP7**	**CASP7**
CCNA2	CDK1	3.401	**CDK1**	**CDK1**
CCNA2	CDK1	3.077	**CDK1**	**CDK1**
PCNA	CDK4	3.045	**PCNA**	**PCNA**
CDK4	PCNA	3.045	**PCNA**	**PCNA**
CASP6	LMNB2	3.164	**CASP7**	**CASP7**
p53	FAS	3.137	**p53**	**FADD**
CCND1	PCNA	3.115	**PCNA**	**PCNA**
Carcinoma				
From (Gene name)	**To (Gene name)**	**Noisy_linkweight**	**Module_source**	**Module_target**
CDK1	CCNB1	3.914	**CDK1**	**CDK1**
CCNB1	CDK1	3.542	**CDK1**	**CDK1**
CCNB1	CDK1	3.542	**CDK1**	**CDK1**
CCNA2	CDK1	3.625	**CDK1**	**CDK1**
CCNA2	CDK1	3.625	**CDK1**	**CDK1**
CDK4	PCNA	3.429	**PCNA**	**PCNA**

PCNA	CDK4	3.429	**PCNA**	**PCNA**
CASP6	LMNB1	3.025	**BAX**	**BAX**
CCND1	PCNA	3.264	**PCNA**	**PCNA**
CASP6	CASP7	3.261	**BAX**	**CASP7**
CDK1	PLK1	2.972	**CDK1**	**CDK1**
LMNB1	LMNB2	2.906	**BAX**	**BAX**
CCNA2	CDK4	3.159	**CDK1**	**PCNA**
HIF1A	Noxa	2.892	**BAX**	**BAX**
GADD45	CDK1	3.120	**CDK1**	**CDK1**
p53	BIK	2.859	**BAX**	**BAX**

Modules highlighted with blue belong to the cell cycle regulation process,
Modules highlighted with red belong to the apoptosis regulation process.

We have listed here only the link weights belonging to the apoptosis or cell cycle related modules. We have found that the number of cell cycle related links is increasing in the adenoma network and even more in the carcinoma network.

Table S8. Representation of the apoptosis and cell cycle related modules among strongest and weakest 10% of the abundances, weighted degrees and the link weights

Table S8. Ratio of cell cycle and apoptosis related modules among the top 10% of mRNA abundances, node degrees and link weights

mRNA abundances

Modules	top 10% normal		top 10% adenoma		top 10% carcinoma		bottom 10% normal		bottom 10% adenoma		bottom 10% carcinoma	
	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%
BAD	14	8.75%	0	0.00%	0	0.00%	9	5.63%	0	0.00%	0	0.00%
CASP3	5	3.13%	0	0.00%	0	0.00%	3	1.88%	0	0.00%	0	0.00%
CDK1	4	2.50%	5	3.13%	4	2.50%	1	0.63%	1	0.63%	1	0.63%
CDK7	1	0.63%	1	0.63%	1	0.63%	0	0.00%	0	0.00%	0	0.00%
FADD	3	1.88%	2	1.25%	4	2.50%	3	1.88%	0	0.00%	1	0.63%
P21	1	0.63%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
BAX	0	0.00%	16	10.00%	22	13.75%	0	0.00%	12	7.50%	10	6.25%
CASP7	0	0.00%	5	3.13%	1	0.63%	0	0.00%	1	0.63%	0	0.00%
CDK4	0	0.00%	3	1.88%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
CDK9	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
CAMP	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
PCNA	0	0.00%	0	0.00%	2	1.25%	0	0.00%	0	0.00%	0	0.00%

Node degrees

Modules	top 10% normal		top 10% adenoma		top 10% carcinoma		bottom 10% normal		bottom 10% adenoma		bottom 10% carcinoma	
	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%
BAD	17	10.63%	0	0.00%	0	0.00%	12	7.50%	0	0.00%	0	0.00%
CASP3	6	3.75%	0	0.00%	0	0.00%	3	1.88%	0	0.00%	0	0.00%
FADD	6	3.75%	5	3.13%	6	3.75%	2	1.25%	2	1.25%	3	1.88%
CDK1	3	1.88%	5	3.13%	5	3.13%	4	2.50%	6	3.75%	5	3.13%
P21	1	0.63%	0	0.00%	0	0.00%	2	1.25%	0	0.00%	0	0.00%
BAX	0	0.00%	19	11.88%	23	14.38%	0	0.00%	15	9.38%	12	7.50%
CASP7	0	0.00%	4	2.50%	1	0.63%	0	0.00%	0	0.00%	0	0.00%

Link weights

Modules	top 10% normal		top 10% adenoma		top 10% carcinoma		bottom 10% normal		bottom 10% adenoma		bottom 10% carcinoma	
	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%
BAD	57	11.40%	0	0.00%	0	0.00%	23	4.60%	0	0.00%	0	0.00%
CASP3	34	6.80%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
CDK1	9	1.80%	23	4.60%	27	5.40%	1	0.20%	1	0.20%	2	0.40%
CDK7	1	0.20%	2	0.40%	2	0.40%	0	0.00%	0	0.00%	0	0.00%

FADD	19	*3.80%*	16	*3.20%*	19	*3.80%*	4	*0.80%*	1	*0.20%*	0	*0.00%*
P21	8	*1.60%*	0	*0.00%*	0	*0.00%*	1	*0.20%*	0	*0.00%*	0	*0.00%*
BAX	0	*0.00%*	83	*16.60%*	90	*18.00%*	0	*0.00%*	20	*4.00%*	15	*3.00%*
CASP7	0	*0.00%*	24	*4.80%*	6	*1.20%*	0	*0.00%*	0	*0.00%*	0	*0.00%*
CDK4	0	*0.00%*	7	*1.40%*	0	*0.00%*	0	*0.00%*	0	*0.00%*	0	*0.00%*
CDK9	0	*0.00%*	0	*0.00%*	0	*0.00%*	0	*0.00%*	0	*0.00%*	0	*0.00%*
PCNA	0	*0.00%*	0	*0.00%*	7	*1.40%*	0	*0.00%*	0	*0.00%*	0	*0.00%*

Table S9. The nodes in the targeted and immunotherapy related pathways

Table S9. Node in targeted and immunotherapy pathways

EGFR signaling				
UniProt	**Name**	**Normal abundance**	**Adenoma abundance**	**Carcinoma abundance**
Q99962	ENDOPHILIN	2.087	1.940	1.607
P01019	AGT	2.397	5.715	6.842
O00291	HIP1	2.831	2.810	4.094
Q14155	P50	3.012	2.924	3.385
P22681	CBL	3.320	3.478	3.375
P30542	A1R	3.360	3.189	2.857
P48023	FASLG	3.715	3.322	3.245
Q05397	FAK	3.965	4.972	5.921
P01133	EGF	4.332	6.502	3.072
O14964	HGS	4.358	4.371	4.951
Q969H0	FBW7	4.825	4.618	5.306
Q99075	HBEGF	4.845	4.958	7.201
P19174	PLCy	4.896	4.982	5.679
P56199	ITGA1	4.942	4.814	10.522
O15530	PDPK1	5.046	4.860	4.399
P42336	PI3K	5.087	4.688	5.842
P18545	PDE6G	5.114	4.935	4.537
Q9P212	PLCE1	5.202	4.869	4.627
P12931	SRC	5.240	5.415	5.119
P15056	bRAF	5.279	5.998	6.407
P27986	P13K	5.390	5.584	4.955
P16591	FER	5.461	5.426	5.892
P78536	ADAM17	5.474	6.019	7.266
P56945	P130Cas	5.798	5.010	4.973
O14944	ER	5.956	4.390	19.968
P01137	TGFB	6.250	5.380	7.047
Q06124	SHP2	6.286	7.684	8.457
Q99704	DOK	6.343	7.151	7.068
P42566	EPS15	6.980	6.085	6.803
Q05209	PTPN12	7.141	9.643	23.956
O75582	MSK1	7.264	6.118	5.671
Q14289	PTK	7.421	6.059	6.103
Q13191	CBLB	7.824	8.672	8.806
O43609	SPRY	7.826	7.182	9.386

Q14451	Grb7	7.969	10.744	11.390
Q9UJM3	MIG6	8.146	8.693	19.414
P00533	EGFR	8.746	8.065	5.623
O75159	SOCS5	8.871	8.789	10.222
Q13480	GAB1	9.354	6.827	5.947
Q8NFH8	REPS2	9.437	15.999	9.841
P62993	GRB2	9.526	8.780	8.831
P31749	AKT	9.871	10.202	10.004
Q07889	SOS1	10.257	7.788	9.171
P51452	M3/6	10.807	10.265	12.322
P00519	ABL1	11.331	10.207	9.253
P05067	APP	11.646	13.139	9.236
P49768	PSEN1	11.843	7.844	6.508
O43639	NCK2	12.192	11.616	11.386
P60953	CDC42	13.341	12.561	11.683
P08913	ALPHA2AR	13.787	9.703	6.168
Q96RT1	ERBIN	14.186	9.313	11.085
P29353	SHC	14.732	13.282	17.058
Q96B97	SH3KBP1	17.578	12.601	12.491
Q9UNH7	SNX6	19.229	17.281	15.822
P46940	IQGAP	20.322	19.421	18.661
Q13882	PTK6	23.078	18.337	17.481
P18085	ARF4	36.423	32.452	34.981
P13688	CEACAM1	123.119	41.697	37.657

VEGFR signaling

UniProt	Name	Normal abundance	Adenoma abundance	Carcinoma abundance
Q15139	PKD	2.663	2.735	3.328
P06241	FYN	2.905	2.368	3.329
P35916	PCLy	3.221	3.231	3.486
P05106	ITGB3	3.320	3.468	3.644
P05771	nPKC	3.324	2.313	2.519
Q15759	MAPK11	3.605	3.460	3.340
Q9UQB8	IRSP53	3.792	4.370	4.134
Q05397	FAK	3.965	4.972	5.921
O75340	Alg-2	3.980	3.669	4.017
Q9ULV1	FZD4	4.122	3.779	4.537
Q14185	DOCK180	4.157	3.946	3.965
O14964	HGS	4.358	4.371	4.951
P18031	PTP1B	4.372	4.393	4.597
P35968	VEGFR	4.523	4.332	7.720
P04004	VITRONECTIN	4.650	4.594	4.353
Q13322	GRB10	4.914	6.006	6.803
P01584	IL1B	5.035	10.157	28.320
P42336	PI3K	5.087	4.688	5.842

P12931	SRC	5.240	5.415	5.119
P27986	P13K	5.390	5.584	4.955
P52735	VAV2	5.523	4.763	5.764
P56945	P130Cas	5.798	5.010	4.973
P15498	VAV	6.185	5.589	5.920
P49137	MAPKAP2	6.639	6.575	7.025
Q15464	EBS	6.662	10.194	10.280
Q14289	PTK	7.421	6.059	6.103
P15692	VEGF	7.524	8.043	12.599
P27540	ARNT	7.571	7.383	6.522
P07900	Hsp90	8.207	10.908	12.175
Q13464	ROCK1	8.858	9.267	8.337
O00459	p85beta	9.284	10.143	9.568
P49023	PXN	9.632	8.813	7.608
P19878	p67phox	9.899	8.391	20.058
Q9Y6W5	WAVE2	9.952	9.746	8.166
P42338	p110la	10.715	8.843	9.766
P04792	HSP27	10.930	8.181	10.793
P16333	NCK	11.062	10.723	10.211
Q16644	MAPKAP-K3	11.250	14.312	14.135
O43639	NCK2	12.192	11.616	11.386
P46108	CRK	12.906	11.606	10.953
O75116	ROCK2	13.225	14.308	12.839
P60953	CDC42	13.341	12.561	11.683
Q16539	p38	14.264	15.622	18.196
Q13177	PAK2	14.496	18.266	17.279
Q8IZP0	E3B1	15.342	17.323	17.234
Q16665	HIF1A	18.121	18.524	27.479
P63000	RAC1	19.531	21.914	23.407
P61586	RHOA	23.631	23.408	24.791
Q9UKW4	VAV3	66.036	54.527	52.007
mismatch repair				
UniProt	**Name**	**Normal abundance**	**Adenoma abundance**	**Carcinoma abundance**
P05129	PRKCG	4.289	4.360	4.045
O15350	P73	4.298	4.486	4.097
Q9Y2T1	AXIN2	5.108	13.637	6.641
Q9NSU2	ATRIP	8.251	8.754	8.721
Q15054	p68	8.815	9.654	9.494
P00519	ABL1	11.331	10.207	9.253
P09429	HMG1	14.850	19.147	20.809
P12004	PCNA	37.079	62.586	71.442

Table S10. Median weighted degrees of targeted and immunotherapy pathway related nodes

Table S10. Median relative weighted degrees of targeted and immunotherapy pathway related nodes

	[a]**Normal -** weighted degree	[b]**Normal -** relative weighted degree	[a]**Adenoma -** weighted degree	[b]**Adenoma -** relative weighted degree	[a]**Carcinoma -** weighted degree	[b]**Carcinoma -** relative weighted degree
EGFR signaling	14.229	*2.592*	14.080	*2.545*	14.207	*2.548*
VEGFR signaling	15.884	*2.894*	15.580	*2.817*	15.651	*2.807*
mismatch repair	5.840	*1.064*	6.063	*1.096*	5.807	*1.042*
the whole network	5.488	*1*	5.532	*1*	5.575	*1*

[a]Weighted degrees are calculated according to the traditional formula, such as the sum of the link weights belonging to each node.

[b]Relative weighted degrees are the quotients of the weighted degrees of each of the signaling pathways and the whole network.

Supporting Codes

Code S1. Calculating diameter for undirected network

```
import networkx as nx
import pdb
import math

# Specify Project Folder
folder = "D:/Bori/diameter/"

filepath = folder + "carcinoma.csv"

#Load all the networks as undirected networks

G = nx.read_weighted_edgelist(path = filepath, delimiter = ';')

#Load all the networks as directed networks

#Gu = nx.read_weighted_edgelist(path = filepath, delimiter = ';')
#G = nx.read_weighted_edgelist(path = filepath, delimiter =
';',create_using=nx.DiGraph())




#These are not complete networks, so this is the removal of the smaller,
isolated parts


"""
#Directed case

Gc = max(nx.connected_components(Gu), key=len)

for component in list(nx.connected_components(Gu)):
    if len(component)<len(Gc):
        for node in component:
            G.remove_node(node)
"""



#Undirected case

Gc = max(nx.connected_components(G), key=len)

for component in list(nx.connected_components(G)):
    if len(component)<len(Gc):
        for node in component:
            G.remove_node(node)
```

```
#Add reciprocal edge value attributes

for u,v,a in G.edges(data=True):
    G[u][v]['reciprocal_weight'] = 1 / G[u][v]['weight']

#Add logarithmic edge weights of normal weights

for u,v,a in G.edges(data=True):
    G[u][v]['logarithmic_weight'] = math.log(G[u][v]['weight'], 10)

#Add inverted edge value attributes

# 1.: find largest edge weight
max = sorted(G.edges(data=True),key= lambda x:
x[2]['weight'],reverse=True)[0][2]['weight']

# 2.: add the inverted weight as an attribute
for u,v,a in G.edges(data=True):
    G[u][v]['inverted_weight'] = max - G[u][v]['weight']

# Do a negative logarithmic mapping of the edge weights

# 1.: Calculate sum of weights for normalizing
# weightSum = G.size('weight')

# 2.: Add the negative logarithmic map weight as an attribute
for u,v,a in G.edges(data=True):
    G[u][v]['normalized_weight'] = G[u][v]['weight'] / max
    G[u][v]['negative_logarithmic_map_weight'] = math.log((G[u][v]['weight'] /
max), 10) * (-1)




    #Shortest path calculation

#With NetworkX shortest path function

path_GNX = nx.shortest_path(G, weight = 'negative_logarithmic_map_weight')
#'weight' for smallest edgeweight diameter; 'reciprocal_weight' for biggest
#path_GNX = nx.shortest_path(G)

#With NetworkX Dijkstra function

path_GD = dict(nx.all_pairs_dijkstra_path(G, weight =
'negative_logarithmic_map_weight')) #'weight' for smallest edgeweight diameter;
'reciprocal_weight' for biggest


#With NetworkX Johnson function
```

```
path_GJ = nx.johnson(G, weight = 'negative_logarithmic_map_weight')
#'weight' for smallest edgeweight diameter; 'reciprocal_weight' for biggest




#Diameters based on just the edge weights

#With NetworkX integrated function

print("\n")
print("Diameters based on just the edge weights:")
print("\n")
print("NetworkX integrated function")
print("\n")

wnx = 0
wnx_greatest = 0
wnxlg = 0
wnxlg_greatest = 0
negativeLogMapWeight = 0
negativeLogMapWeightMax = 0
nodelist = []
shortestPathLengths = []
shortestPathLengthsLg = []
shortestPathLengthsNegativeLogMap = []

for node1, node2 in path_GNX.items():
    for key in node2:
            #print(node2[key])
            for i in range(0, len(node2[key])-1):
                #print(node2[key][i], node2[key][i+1])
                wnx += G[node2[key][i]][node2[key][i+1]]["weight"]
                wnxlg += G[node2[key][i]][node2[key][i+1]]["logarithmic_weight"]
                negativeLogMapWeight +=
G[node2[key][i]][node2[key][i+1]]["negative_logarithmic_map_weight"]
            if negativeLogMapWeight > negativeLogMapWeightMax:
                nodelist = node2[key]
                wnx_greatest = wnx
                wnxlg_greatest = wnxlg
                negativeLogMapWeightMax = negativeLogMapWeight
#             print("Diameter: ", wnx)
            shortestPathLengths.append(wnx)
            wnx = 0
#             print("Diameter (logarithmic): ", wnxlg)
            shortestPathLengthsLg.append(wnxlg)
            wnxlg = 0
#             print("\n")
            shortestPathLengthsNegativeLogMap.append(negativeLogMapWeight)
            negativeLogMapWeight = 0


print(nodelist)
print("\n")
print("Diameter: ", wnx_greatest)
print("Diameter (logarithmic): ", wnxlg_greatest)
print("Diameter (negative logarithmic map): ", negativeLogMapWeightMax)
print("Diameter (edge count): ", len(nodelist)-1)
```

```
print("Average shortest path length: ", sum(shortestPathLengths) /
len(shortestPathLengths))
print("Average shortest path length (logarithmic): ", sum(shortestPathLengthsLg)
/ len(shortestPathLengthsLg))
print("Average shortest path length (negative logarithmic map): ",
sum(shortestPathLengthsNegativeLogMap) / len(shortestPathLengthsNegativeLogMap))


wnx = 0
wnxlg = 0


#With NetworkX Dijkstra function

print("\n")
print("NetworkX Dijkstra function")
print("\n")

wnx = 0
wnx_greatest = 0
wnxlg = 0
wnxlg_greatest = 0
negativeLogMapWeight = 0
negativeLogMapWeightMax = 0
nodelist = []
shortestPathLengths = []
shortestPathLengthsLg = []
shortestPathLengthsNegativeLogMap = []

for node1D, node2D in path_GD.items():
    for keyD in node2D:
            for i in range(0, len(node2D[keyD])-1):
                wnx += G[node2D[keyD][i]][node2D[keyD][i+1]]["weight"]
                wnxlg +=
G[node2D[keyD][i]][node2D[keyD][i+1]]["logarithmic_weight"]
                negativeLogMapWeight +=
G[node2D[keyD][i]][node2D[keyD][i+1]]["negative_logarithmic_map_weight"]
            if negativeLogMapWeight > negativeLogMapWeightMax:
                nodelist = node2D[keyD]
                wnx_greatest = wnx
                wnxlg_greatest = wnxlg
                negativeLogMapWeightMax = negativeLogMapWeight
            shortestPathLengths.append(wnx)
            wnx = 0
            shortestPathLengthsLg.append(wnxlg)
            wnxlg = 0
            shortestPathLengthsNegativeLogMap.append(negativeLogMapWeight)
            negativeLogMapWeight = 0

print(nodelist)
print("\n")
print("Diameter: ", wnx_greatest)
print("Diameter (logarithmic): ", wnxlg_greatest)
print("Diameter (negative logarithmic map): ", negativeLogMapWeightMax)
print("Diameter (edge count): ", len(nodelist)-1)
print("Average shortest path length: ", sum(shortestPathLengths) /
len(shortestPathLengths))
print("Average shortest path length (logarithmic): ", sum(shortestPathLengthsLg)
/ len(shortestPathLengthsLg))
print("Average shortest path length (negative logarithmic map): ",
sum(shortestPathLengthsNegativeLogMap) / len(shortestPathLengthsNegativeLogMap))
```

```
#With NetworkX Johnson function

print("\n")
print("NetworkX Johnson function")
print("\n")

wnx = 0
wnx_greatest = 0
wnxlg = 0
wnxlg_greatest = 0
negativeLogMapWeight = 0
negativeLogMapWeightMax = 0
nodelist = []
shortestPathLengths = []
shortestPathLengthsLg = []
shortestPathLengthsNegativeLogMap = []

for node1J, node2J in path_GJ.items():
    for keyJ in node2J:
            for i in range(0, len(node2J[keyJ])-1):
                wnx += G[node2J[keyJ][i]][node2J[keyJ][i+1]]["weight"]
                wnxlg +=
G[node2J[keyJ][i]][node2J[keyJ][i+1]]["logarithmic_weight"]
                negativeLogMapWeight +=
G[node2J[keyJ][i]][node2J[keyJ][i+1]]["negative_logarithmic_map_weight"]
            if negativeLogMapWeight > negativeLogMapWeightMax:
                nodelist = node2J[keyJ]
                wnx_greatest = wnx
                wnxlg_greatest = wnxlg
                negativeLogMapWeightMax = negativeLogMapWeight
            shortestPathLengths.append(wnx)
            wnx = 0
            shortestPathLengthsLg.append(wnxlg)
            wnxlg = 0
            shortestPathLengthsNegativeLogMap.append(negativeLogMapWeight)
            negativeLogMapWeight = 0

print(nodelist)
print("\n")
print("Diameter: ", wnx_greatest)
print("Diameter (logarithmic): ", wnxlg_greatest)
print("Diameter (negative logarithmic map): ", negativeLogMapWeightMax)
print("Diameter (edge count): ", len(nodelist)-1)
print("Average shortest path length: ", sum(shortestPathLengths) /
len(shortestPathLengths))
print("Average shortest path length (logarithmic): ", sum(shortestPathLengthsLg)
/ len(shortestPathLengthsLg))
print("Average shortest path length (negative logarithmic map): ",
sum(shortestPathLengthsNegativeLogMap) / len(shortestPathLengthsNegativeLogMap))
```

Code S2. Calculating diameter for directed network

```
import networkx as nx
import pdb
import math

# Specify Project Folder
folder = "D:/Bori/diameter/"

filepath = folder + "carcinoma.csv"

#Load all the networks as undirected networks

#G = nx.read_weighted_edgelist(path = filepath, delimiter = ';')

#Load all the networks as directed networks

Gu = nx.read_weighted_edgelist(path = filepath, delimiter = ';')
G = nx.read_weighted_edgelist(path = filepath, delimiter =
';',create_using=nx.DiGraph())




#These are not complete networks, so this is the removal of the smaller,
isolated parts



#Directed case

Gc = max(nx.connected_components(Gu), key=len)

for component in list(nx.connected_components(Gu)):
    if len(component)<len(Gc):
        for node in component:
            G.remove_node(node)



"""
#Undirected case

Gc = max(nx.connected_components(G), key=len)

for component in list(nx.connected_components(G)):
    if len(component)<len(Gc):
        for node in component:
            G.remove_node(node)
"""
```

```
#Add reciprocal edge value attributes

for u,v,a in G.edges(data=True):
    G[u][v]['reciprocal_weight'] = 1 / G[u][v]['weight']

#Add logarithmic edge weights of normal weights

for u,v,a in G.edges(data=True):
    G[u][v]['logarithmic_weight'] = math.log(G[u][v]['weight'], 10)

#Add inverted edge value attributes

# 1.: find largest edge weight
max = sorted(G.edges(data=True),key= lambda x:
x[2]['weight'],reverse=True)[0][2]['weight']

# 2.: add the inverted weight as an attribute
for u,v,a in G.edges(data=True):
    G[u][v]['inverted_weight'] = max - G[u][v]['weight']

# Do a negative logarithmic mapping of the edge weights

# 1.: Calculate sum of weights for normalizing
# weightSum = G.size('weight')

# 2.: Add the negative logarithmic map weight as an attribute
for u,v,a in G.edges(data=True):
    G[u][v]['normalized_weight'] = G[u][v]['weight'] / max
    G[u][v]['negative_logarithmic_map_weight'] = math.log((G[u][v]['weight'] /
max), 10) * (-1)




    #Shortest path calculation

#With NetworkX shortest path function

path_GNX = nx.shortest_path(G, weight = 'negative_logarithmic_map_weight')
#'weight' for smallest edgeweight diameter; 'reciprocal_weight' for biggest
#path_GNX = nx.shortest_path(G)

#With NetworkX Dijkstra function

path_GD = dict(nx.all_pairs_dijkstra_path(G, weight =
'negative_logarithmic_map_weight')) #'weight' for smallest edgeweight diameter;
'reciprocal_weight' for biggest


#With NetworkX Johnson function

path_GJ = nx.johnson(G, weight = 'negative_logarithmic_map_weight')
#'weight' for smallest edgeweight diameter; 'reciprocal_weight' for biggest
```

```
#Diameters based on just the edge weights

#With NetworkX integrated function

print("\n")
print("Diameters based on just the edge weights:")
print("\n")
print("NetworkX integrated function")
print("\n")

wnx = 0
wnx_greatest = 0
wnxlg = 0
wnxlg_greatest = 0
negativeLogMapWeight = 0
negativeLogMapWeightMax = 0
nodelist = []
shortestPathLengths = []
shortestPathLengthsLg = []
shortestPathLengthsNegativeLogMap = []

for node1, node2 in path_GNX.items():
    for key in node2:
            #print(node2[key])
            for i in range(0, len(node2[key])-1):
                #print(node2[key][i], node2[key][i+1])
                wnx += G[node2[key][i]][node2[key][i+1]]["weight"]
                wnxlg += G[node2[key][i]][node2[key][i+1]]["logarithmic_weight"]
                negativeLogMapWeight +=
G[node2[key][i]][node2[key][i+1]]["negative_logarithmic_map_weight"]
            if negativeLogMapWeight > negativeLogMapWeightMax:
                nodelist = node2[key]
                wnx_greatest = wnx
                wnxlg_greatest = wnxlg
                negativeLogMapWeightMax = negativeLogMapWeight
#             print("Diameter: ", wnx)
            shortestPathLengths.append(wnx)
            wnx = 0
#             print("Diameter (logarithmic): ", wnxlg)
            shortestPathLengthsLg.append(wnxlg)
            wnxlg = 0
#             print("\n")
            shortestPathLengthsNegativeLogMap.append(negativeLogMapWeight)
            negativeLogMapWeight = 0


print(nodelist)
print("\n")
print("Diameter: ", wnx_greatest)
print("Diameter (logarithmic): ", wnxlg_greatest)
print("Diameter (negative logarithmic map): ", negativeLogMapWeightMax)
print("Diameter (edge count): ", len(nodelist)-1)
print("Average shortest path length: ", sum(shortestPathLengths) /
len(shortestPathLengths))
print("Average shortest path length (logarithmic): ", sum(shortestPathLengthsLg)
/ len(shortestPathLengthsLg))
```

```
print("Average shortest path length (negative logarithmic map): ",
sum(shortestPathLengthsNegativeLogMap) / len(shortestPathLengthsNegativeLogMap))


wnx = 0
wnxlg = 0


#With NetworkX Dijkstra function

print("\n")
print("NetworkX Dijkstra function")
print("\n")

wnx = 0
wnx_greatest = 0
wnxlg = 0
wnxlg_greatest = 0
negativeLogMapWeight = 0
negativeLogMapWeightMax = 0
nodelist = []
shortestPathLengths = []
shortestPathLengthsLg = []
shortestPathLengthsNegativeLogMap = []

for node1D, node2D in path_GD.items():
    for keyD in node2D:
            for i in range(0, len(node2D[keyD])-1):
                wnx += G[node2D[keyD][i]][node2D[keyD][i+1]]["weight"]
                wnxlg +=
G[node2D[keyD][i]][node2D[keyD][i+1]]["logarithmic_weight"]
                negativeLogMapWeight +=
G[node2D[keyD][i]][node2D[keyD][i+1]]["negative_logarithmic_map_weight"]
            if negativeLogMapWeight > negativeLogMapWeightMax:
                nodelist = node2D[keyD]
                wnx_greatest = wnx
                wnxlg_greatest = wnxlg
                negativeLogMapWeightMax = negativeLogMapWeight
            shortestPathLengths.append(wnx)
            wnx = 0
            shortestPathLengthsLg.append(wnxlg)
            wnxlg = 0
            shortestPathLengthsNegativeLogMap.append(negativeLogMapWeight)
            negativeLogMapWeight = 0

print(nodelist)
print("\n")
print("Diameter: ", wnx_greatest)
print("Diameter (logarithmic): ", wnxlg_greatest)
print("Diameter (negative logarithmic map): ", negativeLogMapWeightMax)
print("Diameter (edge count): ", len(nodelist)-1)
print("Average shortest path length: ", sum(shortestPathLengths) /
len(shortestPathLengths))
print("Average shortest path length (logarithmic): ", sum(shortestPathLengthsLg)
/ len(shortestPathLengthsLg))
print("Average shortest path length (negative logarithmic map): ",
sum(shortestPathLengthsNegativeLogMap) / len(shortestPathLengthsNegativeLogMap))



#With NetworkX Johnson function
```

```
print("\n")
print("NetworkX Johnson function")
print("\n")

wnx = 0
wnx_greatest = 0
wnxlg = 0
wnxlg_greatest = 0
negativeLogMapWeight = 0
negativeLogMapWeightMax = 0
nodelist = []
shortestPathLengths = []
shortestPathLengthsLg = []
shortestPathLengthsNegativeLogMap = []

for node1J, node2J in path_GJ.items():
    for keyJ in node2J:
            for i in range(0, len(node2J[keyJ])-1):
                wnx += G[node2J[keyJ][i]][node2J[keyJ][i+1]]["weight"]
                wnxlg +=
G[node2J[keyJ][i]][node2J[keyJ][i+1]]["logarithmic_weight"]
                negativeLogMapWeight +=
G[node2J[keyJ][i]][node2J[keyJ][i+1]]["negative_logarithmic_map_weight"]
            if negativeLogMapWeight > negativeLogMapWeightMax:
                nodelist = node2J[keyJ]
                wnx_greatest = wnx
                wnxlg_greatest = wnxlg
                negativeLogMapWeightMax = negativeLogMapWeight
            shortestPathLengths.append(wnx)
            wnx = 0
            shortestPathLengthsLg.append(wnxlg)
            wnxlg = 0
            shortestPathLengthsNegativeLogMap.append(negativeLogMapWeight)
            negativeLogMapWeight = 0

print(nodelist)
print("\n")
print("Diameter: ", wnx_greatest)
print("Diameter (logarithmic): ", wnxlg_greatest)
print("Diameter (negative logarithmic map): ", negativeLogMapWeightMax)
print("Diameter (edge count): ", len(nodelist)-1)
print("Average shortest path length: ", sum(shortestPathLengths) /
len(shortestPathLengths))
print("Average shortest path length (logarithmic): ", sum(shortestPathLengthsLg)
/ len(shortestPathLengthsLg))
print("Average shortest path length (negative logarithmic map): ",
sum(shortestPathLengthsNegativeLogMap) / len(shortestPathLengthsNegativeLogMap))
```

Code S3. Calculating diameter for mixed network

```
import networkx as nx
import pandas as pd
import math


def read_create_MultiDiGraph(filename):

    G = nx.MultiDiGraph()

    data = pd.read_csv(filename, delimiter = ';')

    for index, row in data.iterrows():
        if row['Link_Type'] is "0":
            G.add_edge(row['From (Gene name)'], row['To (Gene name)'], weight =
row['Link_Weight'])
            G.add_edge(row['To (Gene name)'], row['From (Gene name)'], weight =
row['Link_Weight'])
        else:
            G.add_edge(row['From (Gene name)'], row['To (Gene name)'], weight =
row['Link_Weight'])
    return G


def remove_isolated_components(g, filename_u):
    G = g

    Gu = nx.read_weighted_edgelist(path = filename_u, delimiter = ';')
    Gc = max(nx.connected_components(Gu), key=len)

    for component in list(nx.connected_components(Gu)):
        if len(component)<len(Gc):
            for node in component:
                G.remove_node(node)
    return G

#Calculate diameter

def diameter(filename, filename_u):

    G = read_create_MultiDiGraph(filename)
    G = remove_isolated_components(G, filename_u)


    #Add reciprocal edge value attributes

    for u,v,a in G.edges(data=True):
        for key in G[u][v].keys():
            G[u][v][key]['reciprocal_weight'] = 1 / G[u][v][key]['weight']

    #Add logarithmic edge weights of normal weights

    for u,v,a in G.edges(data=True):
        for key in G[u][v].keys():
```

```
            G[u][v][key]['logarithmic_weight'] =
math.log(G[u][v][key]['weight'], 10)

    #Add inverted edge value attributes

    # 1.: find largest edge weight
    max = sorted(G.edges(data=True),key= lambda x:
x[2]['weight'],reverse=True)[0][2]['weight']

    # 2.: add the inverted weight as an attribute
    for u,v,a in G.edges(data=True):
        for key in G[u][v].keys():
            G[u][v][key]['inverted_weight'] = max - G[u][v][key]['weight']

    # Do a negative logarithmic mapping of the edge weights
    for u,v,a in G.edges(data=True):
        for key in G[u][v].keys():
            G[u][v][key]['normalized_weight'] = G[u][v][key]['weight'] / max
            G[u][v][key]['negative_logarithmic_map_weight'] =
math.log((G[u][v][key]['weight'] / max), 10) * (-1)



        #Shortest path calculation

    #With NetworkX shortest path function
    path_GNX = nx.shortest_path(G, weight = 'negative_logarithmic_map_weight' )

    #With NetworkX Dijkstra function
    path_GD = dict(nx.all_pairs_dijkstra_path(G, weight =
'negative_logarithmic_map_weight' ))


    #With NetworkX Johnson function
    path_GJ = nx.johnson(G, weight = 'negative_logarithmic_map_weight' )


    #Diameters based on just the edge weights

    #With NetworkX integrated function

    print("\n")
    print("Diameters based on just the edge weights:")
    print("\n")
    print("NetworkX integrated function")
    print("\n")

    wnx = 0
    wnx_greatest = 0
    wnxlg = 0
    wnxlg_greatest = 0
    negativeLogMapWeight = 0
    negativeLogMapWeightMax = 0
    nodelist = []
    shortestPathLengths = []
    shortestPathLengthsLg = []
    shortestPathLengthsNegativeLogMap = []

    for node1, node2 in path_GNX.items():
        for key in node2:
                for i in range(0, len(node2[key])-1):
```

```
                    wnx += G[node2[key][i]][node2[key][i+1]][0]["weight"]
                    wnxlg +=
G[node2[key][i]][node2[key][i+1]][0]["logarithmic_weight"]
                    negativeLogMapWeight +=
G[node2[key][i]][node2[key][i+1]][0]["negative_logarithmic_map_weight"]
                if negativeLogMapWeight > negativeLogMapWeightMax:
                    nodelist = node2[key]
                    wnx_greatest = wnx
                    wnxlg_greatest = wnxlg
                    negativeLogMapWeightMax = negativeLogMapWeight
    #             print("Diameter: ", wnx)
                shortestPathLengths.append(wnx)
                wnx = 0
    #             print("Diameter (logarithmic): ", wnxlg)
                shortestPathLengthsLg.append(wnxlg)
                wnxlg = 0
    #             print("\n")
                shortestPathLengthsNegativeLogMap.append(negativeLogMapWeight)
                negativeLogMapWeight = 0


    print(nodelist)
    print("\n")
    print("Diameter: ", wnx_greatest)
    print("Diameter (logarithmic): ", wnxlg_greatest)
    print("Diameter (negative logarithmic map): ", negativeLogMapWeightMax)
    print("Diameter (edge count): ", len(nodelist)-1)
    print("Average shortest path length: ", sum(shortestPathLengths) /
len(shortestPathLengths))
    print("Average shortest path length (logarithmic): ",
sum(shortestPathLengthsLg) / len(shortestPathLengthsLg))
    print("Average shortest path length (negative logarithmic map): ",
sum(shortestPathLengthsNegativeLogMap) / len(shortestPathLengthsNegativeLogMap))

    wnx = 0
    wnxlg = 0


    #With NetworkX Dijkstra function

    print("\n")
    print("NetworkX Dijkstra function")
    print("\n")

    wnx = 0
    wnx_greatest = 0
    wnxlg = 0
    wnxlg_greatest = 0
    negativeLogMapWeight = 0
    negativeLogMapWeightMax = 0
    nodelist = []
    shortestPathLengths = []
    shortestPathLengthsLg = []
    shortestPathLengthsNegativeLogMap = []

    for node1D, node2D in path_GD.items():
        for keyD in node2D:
                for i in range(0, len(node2D[keyD])-1):
                    wnx += G[node2D[keyD][i]][node2D[keyD][i+1]][0]["weight"]
                    wnxlg +=
G[node2D[keyD][i]][node2D[keyD][i+1]][0]["logarithmic_weight"]
```

```
                    negativeLogMapWeight +=
G[node2D[keyD][i]][node2D[keyD][i+1]][0]["negative_logarithmic_map_weight"]
                if negativeLogMapWeight > negativeLogMapWeightMax:
                    nodelist = node2D[keyD]
                    wnx_greatest = wnx
                    wnxlg_greatest = wnxlg
                    negativeLogMapWeightMax = negativeLogMapWeight
                shortestPathLengths.append(wnx)
                wnx = 0
                shortestPathLengthsLg.append(wnxlg)
                wnxlg = 0
                shortestPathLengthsNegativeLogMap.append(negativeLogMapWeight)
                negativeLogMapWeight = 0

    print(nodelist)
    print("\n")
    print("Diameter: ", wnx_greatest)
    print("Diameter (logarithmic): ", wnxlg_greatest)
    print("Diameter (negative logarithmic map): ", negativeLogMapWeightMax)
    print("Diameter (edge count): ", len(nodelist)-1)
    print("Average shortest path length: ", sum(shortestPathLengths) /
len(shortestPathLengths))
    print("Average shortest path length (logarithmic): ",
sum(shortestPathLengthsLg) / len(shortestPathLengthsLg))
    print("Average shortest path length (negative logarithmic map): ",
sum(shortestPathLengthsNegativeLogMap) / len(shortestPathLengthsNegativeLogMap))



    #With NetworkX Johnson function

    print("\n")
    print("NetworkX Johnson function")
    print("\n")

    wnx = 0
    wnx_greatest = 0
    wnxlg = 0
    wnxlg_greatest = 0
    negativeLogMapWeight = 0
    negativeLogMapWeightMax = 0
    nodelist = []
    shortestPathLengths = []
    shortestPathLengthsLg = []
    shortestPathLengthsNegativeLogMap = []

    for node1J, node2J in path_GJ.items():
        for keyJ in node2J:
                for i in range(0, len(node2J[keyJ])-1):
                    wnx += G[node2J[keyJ][i]][node2J[keyJ][i+1]][0]["weight"]
                    wnxlg +=
G[node2J[keyJ][i]][node2J[keyJ][i+1]][0]["logarithmic_weight"]
                    negativeLogMapWeight +=
G[node2J[keyJ][i]][node2J[keyJ][i+1]][0]["negative_logarithmic_map_weight"]
                if negativeLogMapWeight > negativeLogMapWeightMax:
                    nodelist = node2J[keyJ]
                    wnx_greatest = wnx
                    wnxlg_greatest = wnxlg
                    negativeLogMapWeightMax = negativeLogMapWeight
                shortestPathLengths.append(wnx)
                wnx = 0
```

```
                shortestPathLengthsLg.append(wnxlg)
                wnxlg = 0
                shortestPathLengthsNegativeLogMap.append(negativeLogMapWeight)
                negativeLogMapWeight = 0

    print(nodelist)
    print("\n")
    print("Diameter: ", wnx_greatest)
    print("Diameter (logarithmic): ", wnxlg_greatest)
    print("Diameter (negative logarithmic map): ", negativeLogMapWeightMax)
    print("Diameter (edge count): ", len(nodelist)-1)
    print("Average shortest path length: ", sum(shortestPathLengths) /
len(shortestPathLengths))
    print("Average shortest path length (logarithmic): ",
sum(shortestPathLengthsLg) / len(shortestPathLengthsLg))
    print("Average shortest path length (negative logarithmic map): ",
sum(shortestPathLengthsNegativeLogMap) / len(shortestPathLengthsNegativeLogMap))



diameter('carcinoma.csv', 'carcinoma_u.csv')
```