Krishna Kanhaiya | Eugen Chzeler | Cristian Gratie | Ion Petre

Controlling Directed Protein Interaction Networks in Cancer

Turku Centre *for* Computer Science

TUCS Technical Report
No 1155, April 2016

Controlling Directed Protein Interaction Networks in Cancer

Krishna Kanhaiya
Computational Biomodeling Laboratory
Turku Centre for Computer Science
Åbo Akademi University, 20540 Turku, Finland
kkanhaiy@abo.fi
Eugen Chzeler
Computational Biomodeling Laboratory
Turku Centre for Computer Science
Åbo Akademi University, 20540 Turku, Finland
eczeizle@abo.fi
Cristian Gratie
Computational Biomodeling Laboratory
Turku Centre for Computer Science
Åbo Akademi University, 20540 Turku, Finland
cgratie@gmail.com
Ion Petre
Computational Biomodeling Laboratory
Turku Centre for Computer Science
Åbo Akademi University, 20520 Turku, Finland
ipetre@abo.fi

TUCS Technical Report
No 1155, April 2016

Abstract

Control theory is a well-established approach in network science, with known applications in bio-medicine and cancer research. We build on recent results for full and structural controllability of directed networks, which gives a set of driver nodes able to control the whole network, or an a-priori defined part of it, respectively. We develop a novel approach for the structural controllability of cancer networks and demonstrate it for the analysis of breast, pancreatic, and ovarian cancer. We build in each case a signalling transduction (STN) protein-protein interaction (PPI) network and focus on the so-called "essential proteins" specific to each cancer type in our study. We show that the cancer essential proteins are efficiently controllable from a (relatively small) computable set of driver nodes. Moreover, we adjust the method to find the driver nodes among FDA-approved drug-target nodes. Interestingly, we find that while many of the drugs acting on our driver nodes are part of known cancer therapies, some of them are not used for the cancer types analyzed here; also some drug-target driver nodes identified by our algorithms are not known to be used in any cancer therapy. Overall we show that a better understanding of the control dynamics of cancer through mathematical modelling could pave the way for new efficient therapeutic approaches and personalized medicine.

Author Summary

Advances in systems biology are offering not only insight into complex molecular interactions, but are also useful for the discovery of new disease proteins and of new therapeutic targets for disease intervention. Here we employ a control theory approach for the analysis of specific disease networks, allowing us to drive the system dynamics towards favourable traits, as well as helping us to understand better the regulatory mechanisms of these bio-chemical networks. We show how to employ the use of well established drug-target proteins in order to archive a structural control over essential target proteins within specific cancer protein-protein interaction networks, and apply this to breast, pancreatic, and ovarian cancer signalling transductions PPIs networks. We demonstrate that instead of aiming for an overall control of entire networks, partial controllability is more effective and efficient in the development of therapies for various cancers. Also, we provide a new insight into the efficient control of dynamical disease networks which can assist in the discovery of novel cancer associated proteins and bio-markers.

Introduction

The main cause of cancer is genetic and epigenetic alterations, which allow normal cells to over-proliferate as tumour cells [1]. Most of these alterations contribute to various cancer dysregulated signal transduction pathways, which control essential cell processes such as growth factor, differentiation and survival [1]. Through signal transduction processes, these tumour cells develop as malignant cells; this complex process information is transmitted through protein-protein interactions (PPIs) [1]. Proteins act as the vehicles of these signals, while the interactions among them influence the velocity of the information flow. For instance, PPIs are directly regulating the phosphorylation of serine/threonine residues [2], and the same process is used by tumour necrosis factor to convey signals from the receptor to their downstream targets [2]. Also, the transforming growth factor-β (TGFβ) employs PPIs to convey signals to activate its targets [2]. TGFβ interacts also with other signaling pathways [3] and creates a complex web in cancer signaling. It has been shown that TGFβ also regulates various kinase cascades such as the mitogen-activated protein kinase (MAPKs) ERK, the p38 MAPK pathways, the Jun N-terminal kinase (JNK), the PI3K kinase, the PP2A phosphatases and the Rho family members [4, 5]. Furthermore, by using docking proteins and protein interaction domains, the receptor tyrosine kinases (RTKs) recruits targets to the receptor [2]. These protein domains mediate a series of intra-molecular interactions during the downstream of RTKs and re-wire the signaling networks [6]. Usually, RTK modules are highly mutated and over-expressed in cancer, which effectively leads its signaling to escalate the progression of tumours [7]. Also, RTKs help to build robust cancerous signalling networks and signal to other tumor cells to form similar networks [8]. These studies show that to comprehensively understand the specificity in signaling networks, we have to understand how distinct pathways communicate with each other and how proteins of one pathway make interactions with related signaling components. A network approach over the cancer's signal transduction dynamics gives us the tools to provide a better understanding of the various information-processing abilities employed during the molecular alteration of the cancerous cells [9].

In human diseases, both associated and non-associated diseased proteins interact with one-another to create disease modules [10], and pave the way towards a layered configuration and understanding of these complex diseases [11]. Previous studies have shown that networks associated with the same disease family, as well as with common phenotypes tend to contain significant similarities between their disease modules [12]. Disease proteins produce some common tendencies, such as: inside and outside interactions of modules through PPIs, co-expresed in specific tissues, as well as high expression correlations [13]. Uncovering these disease-specific interactions is essential not only in demonstrating the complex molecular mechanism inside these networks, but also in providing an inside-view of the dysfunctional signaling transduction processes within these networks. All these examples illustrate that a network approach toward disease analysis could provide significant new insights into disease-gene identifications, as well as it could open new approaches towards network-based therapeutic tools, targeting entire disease modules together instead of individual elements [14]. The current system-based understanding of biological processes has already showed that due to the various overlaps of signaling pathways, proteins participating in multiple pathways build robust inter-pathways connections. Therefore multi-target drugs can inhibit multiple proteins and can thus increase the chance of effective treatments [15]. In turn, targeting single protein can damage the connection of multi-cellular functions and delay the recovery of disease [15, 16, 17]. However, in many diseases, the relationship between the various drug targets and the associated disease proteins is still vague. This opens a new door of investigation for finding rational disease control mechanisms by use of the currently available drug-target proteins. Essential proteins are of the central interest in such investigations, in identifying novel targets for therapeutics [18]; there is already evidence that targeting essential proteins in cancer can lead to novel therapies [19, 20]. Proteins are consider essential in cancer if their mutations cause the death of the cancer cells [21]. Cancer essential proteins can be found in specific cell lines and often induce oncogenesis [20].

Network biology, with the help of mathematical modeling, has revolutionized the human diseasome research and paved the way towards the development of new therapeutic approaches and personalized medicine [22]. This is why, in the last couple of decades, networks science has been constantly in the focus of biological research, where scientists try to understand the dynamics and control features of various complex bio-chemical networks in association with matching experimental findings [23]. Recent work on network controllability has shown that full controllability and reprogramming of inter-cellular networks, which assumes the driving of the complete network from any initial state to any desired final state, can be achieved by a minimum number of control targets [22, 24]. However, the computer-based experimental tests of Liu et al [25] suggest that achieving full control over gene regulatory networks is rather demanding, requiring sometimes up to 80% of the nodes to be directly controlled by an external controller. At the same time, another research by Wuchty et al. [26] demonstrated the existence of the so called minimum dominating sets (MDSets) of proteins inside protein interaction networks. These MDSets are groups of proteins which have a central position within the protein interaction networks, and are interacting with all other proteins within the network. They also showed that the MDSets are enriched with essential, cancer-related and virus-targeted proteins, which are acting as bottlenecks for various essential cell processes. Based on the study from [26] and considering essential MDSets (e-MDSets), Khuri et al. [27] showed that e-MDSet proteins have predominantly more connections in networks than any other sets of proteins, and can be vital for network control. Another framework based on feedback loops (both negative and positive) showed that these loops play a major role in signaling transduction networks by causing various oscillations and switching of signals. Thus, they concluded that also these feedback loops could be a major target for controlling oncoproteins and for the development of effective therapies in cancer [23]. All the above approaches are aimed at controlling specific types of biological networks. However, in the case of diseased protein-protein interaction networks, there is still no feasible methodology for achieving an efficient control.

In this article we use target controllability for the analysis of specific signal transduction cancer networks, focusing on cancer type specific essential proteins as our target nodes and on drug-target proteins as our driven nodes. We develop a general computational model based on directed networks, that aims to find specific paths from the set of potential driven nodes to the set of targets. We report on the total number of driven nodes needed to control the targets, the number/list of drug targetable driven nodes, and on some interesting topological properties of the driven nodes in these networks.

Materials and Methods

Cancer data

The cancer data used in this study was obtained both from the publicly available UniprotKB protein database [28], as well as from previous published research articles [29, 30, 31, 32, 33, 34, 35, 36]. We concentrated our study over three types of cancer, namely breast, pancreatic and ovarian, for which we gathered data for 1415, 991 and 1047 proteins respectively, see S1 Table. While gathering this data we required that the selected proteins should be reviewed in curated databases and/or gathered from literature. We used short python scripts to check for redundancy in the gathered data.

Protein-protein interaction data

To obtained directed PPI cancer data, we used SIGNOR (SIGnaling Network Open Resource) database [37], which outputs binary matrix representations for the used-provided protein lists; this allowed us to create directed graphs between signaling entities. We obtained directed PPIs network of 2532 interactions from 1415 nodes in breast cancer, 1569 interactions from 991 nodes in pancreatic caner, and 1643 interactions from 1047 nodes in ovarian cancer. The networks are available at [38] or see S2 Table

Essential protein data

Although diseased cells may harbor hundreds of genomic alterations in various biological pathways [9, 24], only a subset of these alterations are driving the disease initiation and progression. These proteins form together the sets of (disease specific) essential proteins. Due to the new CRISPR gene editing technology, researchers can now pinpoint essential proteins for a very large class of illnesses [39], including many types of cancers [40, 41]. We collected essential gene data for all three types of cancer from the COLT-Cancer database [42]. In particular, we considered the MDA-MBD-231, HPAF-II and OV-90 cell lines respectively for breast, pancreatic and ovarian cancer, and follow the GARP (Gene Activity Rank Profile) and GARP-P value of corresponding proteins mentioned in the database. Since previous studies showed that proteins with lower GARP score are more essential and directly associated with oncogenesis [40], we selected only those essential proteins whose GARP value is in the negative range, and moreover, whose GARP-P value is less than 0.05 ($p \leq 0.05$). Following the above criteria, we identified 712, 770 and 866 proteins respectively for breast, pancreatic and ovarian cancer, see S3 Table. Out of these, 135, 168 and 140 essential proteins respectively in breast, pancreatic and ovarian cancer were found available in the SIGNOR PPI network database, and were included in our network.

Drug target data

We obtained drug-target protein data from the open source DrugBank database [43]†. The DrugBank database offers extensive information of drug and drug targets. This includes information of chemical, pharmacological and pharmaceutical specific drugs integrated with structure, pathway and sequence drug target. For drug-target identifiers we have selected in total 1507 FDA-approved proteins which have a known mechanism, see S4 Table.

Theoretical model and optimization algorithm

The mathematical methodology used for deriving the sets of driving genes through which we can effectively manipulate the system is based on the well established Structural Control Theory. This theory, although initiated by Lin [44] in the 70's, has recently received a new boost of attention, [22, 25, 45] partly due to recent results on efficient algorithms for core research problems within this framework.

We say that a dynamical system, such as the expression levels of a set of genes influencing each other, is *controllable* from a set of input (*driver*) nodes, if there exists a time-dependent sequence of input signals delivered through these nodes such that the system can be driven from any initial state to any desired final state within finite time. From the point of view of our study, we can concentrate over linear time-invariant (lti) dynamical systems. Such systems can be visualized as directed networks, where the nodes represent the components of the system while the weighted directed edges represent how these components interact and influence each other.

In a recent breakthrough an efficient (low polynomial time) algorithm was provided for computing the minimal number of input nodes needed to structurally control any given lti network [25]. However, it was also shown that in the case of sparse inhomogeneous networks, such as most of the networks emerging from biochemical and biomedical applications, controlling the entire system is expensive,

†The query was performed on August 2015

requiring up to 80% of the system's nodes to be controlled directly. On the other hand, in terms of practical applications, in many cases it is enough to control only a certain well-selected portion of the network's nodes, such as the set of essential genes, in order to impose a certain overall behaviour over the system. Thus, controlling those target genes, or a considerable subset of them, could translate into a highly effective control approach over the desired system dynamics.

Our algorithms aim to minimize the number of driven nodes (i.e., network nodes) which can be used to control a given target, namely the set of cancer-specific essential genes in each network. Our approach is different that in [22] that minimize the number of input/driver nodes (i.e., possibly acting upon several of the network nodes in the same time). The rationale for this choice is that we aim for combinatorial drug target identification and we consider only the primary target of each drug under consideration. Our algorithm has a double optimization to minimize the total number of driven nodes (on which a subsequent intervention is needed) and to maximize the percentage of FDA-approved target nodes among them. We used several heuristic strategies for a more efficient exploration of the search space, leading to faster and better results.

We implemented an additional validation step for the proposed solution of our algorithm, which is freely available at [38]. An example in [46] shows that in some rare cases, the algorithm in [22], whose basic search strategy we also follow here, may output a non-solution (a set of nodes that fails to control the given target). In such a case, we restart the search algorithm and given the built-in randomness of our algorithm, we expect to get another candidate solution with high probability. The size of the set of non-solution output is not-known in general but according to [36], it is expected to be very small. This is consistent with our computational results where no non-solutions were found.

Topological properties of networks

The *degree* of a node in a network is the number of connections the node has to other nodes. The robustness of a network depends upon the connections between the nodes inside the network. Another important node-associated value is the *cluster coefficient*, which, for a node v, is defined as $C_v = n/k_v(k_v - 1)$, where k_v is the number of neighbours of v and n is the total number of connections/edges between these neighbours. The clustering coefficient of a node takes values between 0 and 1, where 1 implies that the node v is in a complete sub-graph, while 0 denotes that the node is part of a loosely connected cluster (a star-shaped cluster with v in the centre). Further, the *betweenness centrality* of a node v is defined as the weighted sum of all shortest pathes between all pairs of nodes s and t, that go through the node v. That is, $C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$, where σ_{st} is the number of shortest paths between nodes s and t, while $\sigma_{st}(v)$ is the number of such shortest paths running through node v. Also, the *closeness centrality* of a node indicates how close this node is from all other nodes; it is defined formally as $C_c(v) = \sum_{t \neq v \in V} \frac{S(s,t)}{n-1}$, where $S(v, t)$ is the shortest path between v and t.

Results

Controlling PPI signaling transduction networks in cancer

To determine the controllability of essential proteins we analyzed three cancer networks: breast, pancreatic and ovarian cancers, see S2 Table. The size of the PPI networks that we generated based on [33] ranges roughly between 900-1600 nodes and between 1500-2500 edges, Table 1.

We first computed the minimum set of nodes controlling the entire network, based on the algorithm for full controllability in [25]. We found in all three cases that around 70% of all the nodes have to be directly controlled in order to gain control over the whole network, Table 1. This is in accordance with previous results of [25] for different types of gene regulatory networks and confirms that full controllability is impractical in cancer medicine.

We then considered a set of cancer type-specific essential proteins based on [42] and computed the sets of nodes that are enough to (target) control these essential proteins, see S3 Table. We applied both the generic algorithms of [22] (whose search we improved through a new heuristic strategy) and our algorithm maximizing the use of drug-targetable nodes as driven nodes, aiming to make the results more practical. The results are summarized in Table 2; the cancer PPI networks are graphically described in Fig. 2 for pancreatic cancer, S1 Fig for breast cancer, and S2 Fig for ovarian cancer. We found that the number of driven nodes needed for the control of essential proteins is much smaller, ranging between 6-14% of the total number of nodes in the network, Fig. 1, depending on the cancer type and on the algorithms used in the computation. Our algorithms also found sets of driven nodes containing many (19-32) drug targetable nodes, Table 2, drastically improving the applicability of this approach.

Table 1: **Full controllability of three cancer network**

Network	**Nodes**	**Edges**	**Full control: driver (drug-targ.) nodes**	**Full control: % driver (% drug-targ.) nodes**
Breast	1415	2532	962 (210)	68% (15%)
Pancreatic	991	1569	690 (472)	70% (48%)
Ovarian	1047	1643	736 (432)	70% (41%)

The columns represent the following information per cancer network: the total number of nodes in the network (Nodes), the number of connections (Edges), the controlling set for the entire network (Full control), the % of the controlling set vs. the whole network (Full control %).

Table 2: **Essential gene -targeted controllability of three cancer networks**

Network	**Nodes**	**Edges**	**Targets: #(%)**	**Min target control: driven (drug_tar)**	**Min target control(%): driven (drug_tar)**	**Drug_oriented target control: driven (drug_tar)**	**Drug_oriented target control(%): driven (drug_tar)**
Breast	1415	2532	135 (9.5%)	94(1)	6.6% (0%)	110(19)	7.7% (1.3%)
Pancreatic	991	1569	168 (17%)	131(9)	13.2% (0.9%)	143(32)	14.4% (3.2%)
Ovarian	1047	1643	140 (13%)	111(6)	10.6% (0.5%)	120(25)	11.4% (2.3%)

The columns represent the following information per cancer network: the total number of nodes in the network (Nodes), the number of connections (Edges), the number (and percentage) of target proteins (Targets), the minimal controlling set of the target proteins (Min target control) including enclosed drug-target (drug_tar) proteins, percentage (vs. the whole network) of the minimal controlling set (Min target control (%)) including enclosed drug-targets, the drug-oriented controlling set (Drug-oriented target control) including enclosed drug-target (drug_tar) proteins, and the percentage (vs. the whole network) of the drug-oriented controlling set (Drug-oriented target control(%)) including enclosed drug-targets (drug_tar).

These results portray a highly advantageous situation. The ultimate objective of our use of structural controllability on cancer disease networks is to be able to control the cancer evolution and drive it towards a downfall. This can be achieved by gaining control over the entire network, an approach which seems to require an excessive direct intervention over 68%, 69.6%, and 70.3% of the nodes in the network, i.e., 962, 690, and 736 of the nodes from the breast, pancreatic and ovarian cancer networks, respectively, Fig. 1 and Table 1. In contrast to the previous situation, we aimed in achieving a control over a subset of nodes, specific to each individual cancer network, which is known to have an overwhelming effect over that cancers survivability. The reasoning is that by controlling these essential positions in the network, we ensure the overall down-evolution of the disease. In contrast to the previous case, controlling these focused target can be done much more efficiently, requiring a direct intervention over 6.6% (94 nodes) 13.2% (131 nodes) and 10.6% (111 nodes) of the breast, pancreatic and ovarian cancer networks, respectively. Thus, we obtain up to a 10 fold decrease in the control effort, Fig. 1, while maintaining a high likelihood of an overall similar effect.

Topological properties of drug target proteins and of essential proteins

We analyzed several topological properties of the drug target proteins included by our algorithm in the set of driven nodes, and of the essential proteins in each of the networks in our study. We looked at the average degree, the betweenness centrality, the closeness centrality, and the clustering coefficient of these proteins as compared

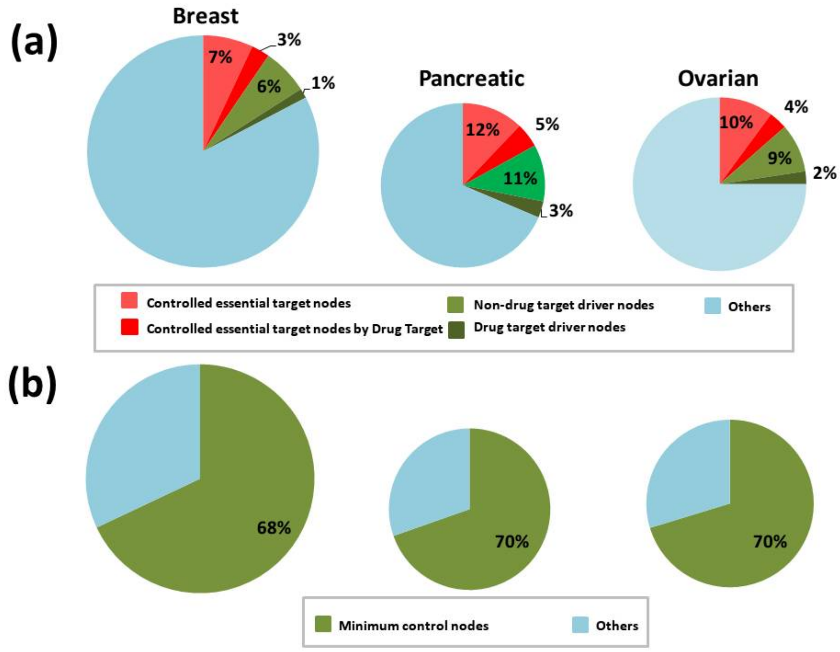

Figure 1: **Controlling of cancer networks.** The radius of the circles is proportional with the number of nodes in the networks. (a) The percentage of controlled target nodes by drug-target nodes and non drug-target nodes, w.r.t. the total number of nodes. (b) Required minimum control nodes for the control of the whole cancer networks.

with the average values over the entire networks. We found that in all the three considered cancer networks, the drug-target driven nodes and the essential proteins have much higher average degree than the average over the whole networks, Fig. 3. This shows that both the drug-target driven nodes and the essential proteins are hubs in the networks and thus central in the regulation of the networks; this is consistent with observation of, e.g., [47, 48]. The essential proteins were found to have a higher average betweenness centrality than the average over the whole networks, especially in the breast and in the pancreatic cancer networks, Fig. 3. This indicates that essential proteins act as highly-traversed bridges in these interaction networks; nodes with high betweenness centrality values have been reported also in several other pathways, including MAPK pathways [49, 50].

The other topological indicates we considered did not systematically distinguish the drug-target driven nodes or the essential proteins against the rest of the networks nodes, Fig. 3.

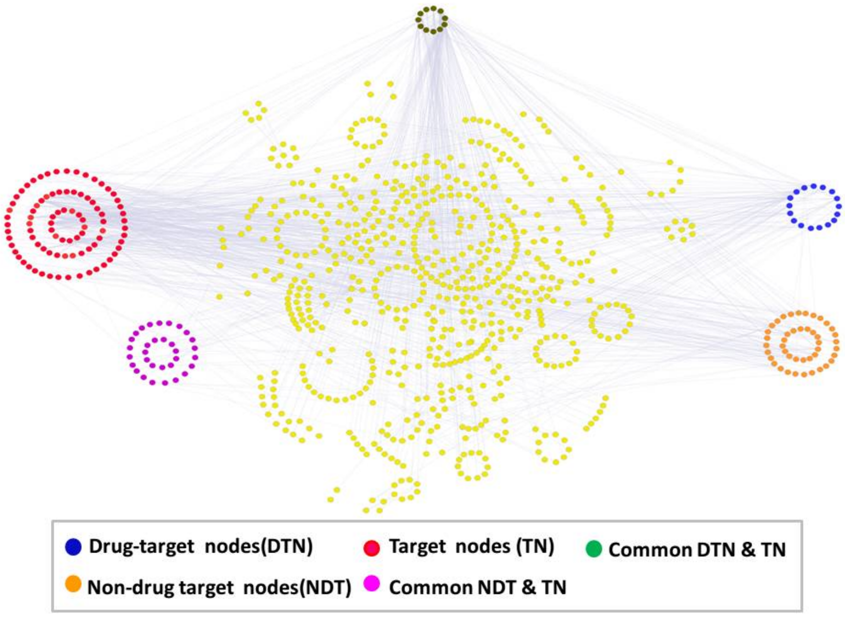

Figure 2: **Pancreatic cancer PPI network.** A yeast two-hybrid (Y2H) PPI network for cancer signalling proteins. A network view of pancreatic cancer PPI network. The network contains 90% of the total network nodes, the remaining part of the network containing isolated nodes. The drug-target nodes (DTN) are shown in dark blue, target nodes (TN) are in maroon, nodes that are both in DTN and TN are shown in dark green, Non-drug target nodes (NDT) are in light orange, and nodes that both in NDT and TN are in magenta.

From driven nodes to combinatorial drug therapy strategies

We analyzed the drug-targetable proteins identified by our algorithms as part of the strategies to control the cancer essential proteins. We found that some of them are themselves oncoproteins and thus could be a direct target in cancer therapy. Among those that are not oncoproteins, we found that some have a high impact in their corresponding network, controlling several essential proteins simultaneously. One of them is ERBB2, which controls five essential proteins in breast cancer (CDK1, CDC27, CDC7, SH3RF1, APLP2) and four essential proteins in pancreatic cancer (CNSK1E, MST1R, MAML1, ADAM17); Fig. 4 and Table 3. This is in line with previous observation of [51] showing that ERBB2 is often a drug-target in cancer therapies. Another potent drug-target protein is RET, controlling five oncoproteins (MAPK3, PLK1, OPTN, PTTG1, CDH1) in the ovarian cancer network, Fig. 4 and Table 3. The list of all high impact drug-target protein (controlling more than two essential proteins) is in Table 3. We observed that out of the 75 drug-target proteins included by our algorithms in the control strategies (driven nodes) of the three

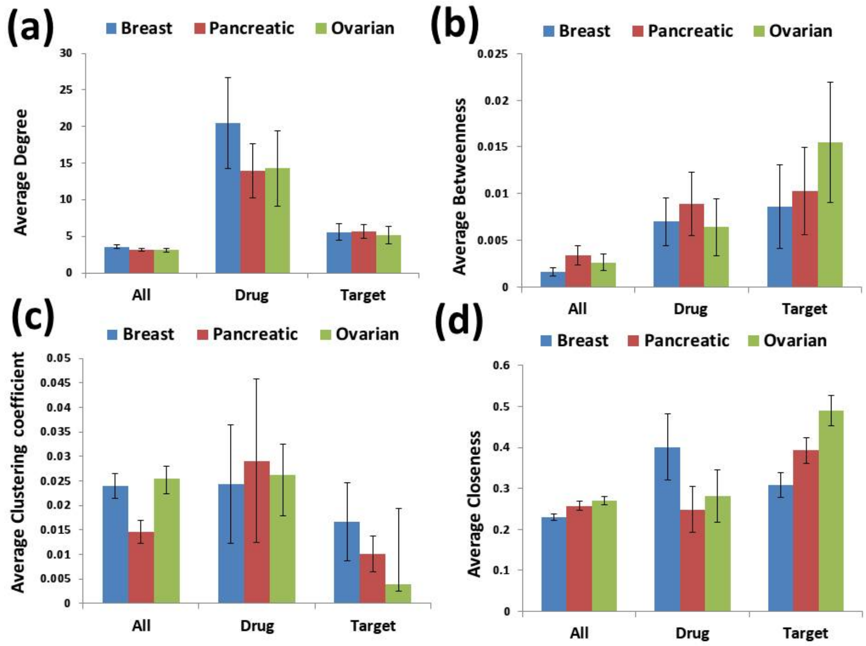

Figure 3: **Topological properties of drug-target and target (essential) proteins in compare to whole network.** **(a)** Average degree of drug-target and target proteins in compare to whole network. **(b)** Average betweenness of drug-target and target proteins in compare to whole network. **(c)** Average clustering coefficient of drug-target and target proteins in compare to whole network. **(d)** Average closeness of drug-target and target proteins in compare to whole network.

cancer networks, 31 of them are present in more than one cancer network. This shows that they are expressed in multiple cancer cell lines and could be used in drug therapies of several cancer, in combinations with cancer type-specific targets.

We looked for anti-cancer drugs for the drug target proteins identified by our control algorithms. We found that in some cases they are used in current cancer type-specific drugs and drug-therapies; for example, anti-cancer drugs targeting the ERBB2 gene are in use for breast cancer. In many other cases however, we found that the drug-targets identified by our methods are either not used in any known cancer therapies, or at least not in the case of the specific cancers we analyzed. These results and observations are summarized in, Table 3.

Functional properties of the high-control proteins

We reviewed the functional properties of the driven nodes found to have the highest impact in controlling the essential proteins; they are ERBB2, SRC, PDPK1, PRKDC, mTOR for breast cancer, ERBB2, AKT1, GSK3B, ABL1 in pancreatic cancer, and RET in ovarian cancer. Our goal was to correlate the findings of our

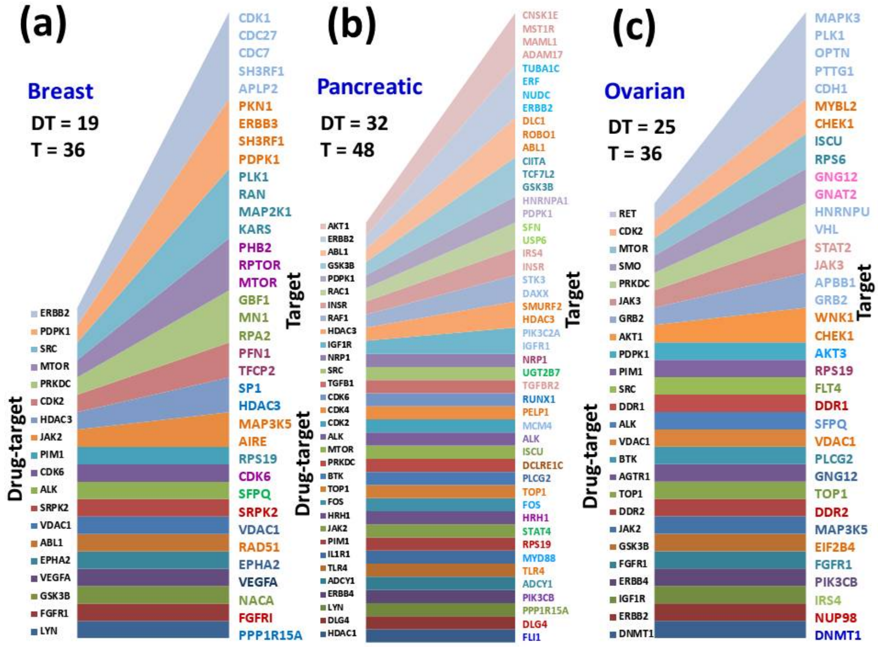

Figure 4: **Target control efficiency of drug-target proteins.** **(a)** Control features of drug-target proteins in breast cancer. **(b)** Control features of drug-target proteins in pancreatic cancer. **(c)** Control features of drug-target proteins in ovarian cancer.

computational study with previous studies on the functional properties of these proteins.

In breast cancer, the ERBB2 oncogene activates signaling pathways that deregulate the essential proteins processes and make cancer cells resistant to chemotherapeutic drugs of cancer cells [52]. Amplification of ERBB2 gene is the main cause of its over-expression in cancer [53], while depletion of glucose also inhibits expression of ERBB2 [52]. Moreover, the PI3K/AKT pathway directly activates the mutations in ERBB2-amplified breast cancers [54]. The SRC protein is activated by various factors such as cytoplasmic proteins, which play vital role in integrating signalling and ligand activation of cell surface receptors. These interactions interrupt the intermolecular interaction within SRC and lead to over-expression of upstream growth factor receptors [55]. Other intrinsic factors in breast cancer are dephosphorlization of SRC, SRC regulation by RTKs, and SRC activity gene expression signature [56].

The protein PDPK1 has a crucial role in over-proliferation of breast cancer [57], while [58] shows that anchorage-independent growth is regulated by PDPK1, which resists to many anti-cancer drugs and starts the tumour formation in breast cancer cell lines. Along with this, PDPK1 proteins phosphorylate the activating segment of AKT, which affects various key cell functions and facilitate the breast cancer progression [57]. The PRKDC protein is also carrying an important role in breast cancer [59, 60]. Downregulation of MYC mRNA and protein expression in multiple cancer cell lines is caused by inhibition of PRKDC, which leads to over-expression of MYC family of proteins induced DNA double-strand breaks and leads to cancer progression [61]. Protein mTOR, together with PIKS and Akt, mediates multiple cellular pathway functions. Aberrations and degradations inside these pathways leads to tumour proliferation in breast cancer [59, 62]. These aberrations affect germline and somatic mutations, amplification, rearrangements, methylation, overexpression, aberrant splicing, and starts mutation in breast cancer cell lines [64].

In pancreatic cell lines, overexpression of ERBB2 is known to advance the disease states [64]. Moreover, knocking down of CAPAN-1 and CAPAN-2 cells by ERBB2 increases the sensitivity to gemcitabine, the resistance to irinotecan/SN-38, the increase of hCNT1 and hCNT3 transporters, and ABCG2, MRP1 and MRP2 ATP-binding cassette transporters expression, which leads to apoptosis [65]. In vivo, PEAK1-dependent kindles induced by oncogenic KRas amplify the loop between SRC, PEAK1, and ERBB2 drive pancreatic cancer. Also, increased SRC-dependent PEAK1 expression by blockade of ERBB2 expression activates tumour growth [66]. The next protein in our list is AKT1, which is serine/threonine kinase AKT (also known as Protein Kinase B) for which we found reports of over-expression in pancreatic tumour formation [67]. The alteration of AKT increases the oncogenic changes in tumour, and the activation of AKT isoforms disturbs the down-regulation of pancreatic tumours which starts upstream signalling [67]. Also, the activation of HER2/3- PI3K/Akt signaling pathways by VIP plays a key role in growth and survival of cancer [68]. For the next protein in our list GSK3B, we found reports that its inhibition activates JNK-cJUN-dependent apoptosis in human pancreatic cancer cell lines [69] and participates in the nuclear $factor-k\beta$ $(NF-k\beta)$ mediated cell survival in pancreatic cancer [70]. Also, GSK3B is documented to initiate the tumour through activation of the oncogenic β-catenin [71], which over-expressed the GSK3B in pancreatic cancer. The next protein ABL1 is over-expressed in pancreatic cancer [72]. Alteration in ABL mRNA expression in tumours increases the activity of ABL kinase, which promotes the cancerous' cell over-proliferation and survival [73]. Interestingly, cellular stress and DNA damage induced ABL1 escalate the cell growth arrests or apoptosis mediated by p53 or p73 [74].

In ovarian cancer RET (REarranged during transfection) is expressed and involved in pathogenesis of ovarian cancer [75, 76]. RET tyrosine kinase is a fusion partner of TRIM27 (tripartite motif-containing 27), which is highly expressed in normal epithelial cells of the ovary and fallopian tube and in ovarian serous carcinoma cells [77]. It has been pathologically characterized in patients with ovarian serous carcinoma. Since RET participates in essential cellular processes, the over-expression of fusion proteins (TRIM27-RET) disrupts its essential cellular activity and triggers tumorigenesis [77].

Discussion

We analyzed the breast, pancreatic, and ovarian cancer protein-protein interaction networks, and identified the respective sets of driven proteins for controlling the networks. Recent genetic editing technologies explain the existence of cancer-specific sets of proteins which have an important role in the overall disease mechanism. These proteins, called cancer essential proteins, are proved to be key for *in-vivo* cancerous cell's proliferation and survival. Therefore, instead of trying to achieve a full control of the entire disease's network, which in itself is highly complex, our approach aims for a targeted control approach, particularly for controlling those cancer essential proteins. In order to achieve the partial control of all the above mentioned cancers, we have first generated for each of them the associated signal transduction directed protein-protein interaction network. These networks identify the in-between influence of the proteins passed on their overall expression levels. Our analysis showed that in order to control all of the essential proteins in these cancer networks, we require the direct intervention over only 6.6% - 13% of the entire networks' nodes, Table 2. In turn, to achieve a full control of these networks, it required around 70% of the networks' nodes to be directly controlled by an outside intervention, Table 1, e.g., such as achieved by administering a number of drugs. Thus, our method generates up to a 10-fold decrease in the control effort, while maintaining a high likelihood of an overall similar effect. Moreover, our methodology and algorithms for target control of the essential proteins emphasize, and maximize, the use of known drug target proteins, as a choice for input controlling nodes, i.e., driven nodes, of the network. Our analysis shows that when comparing the drug target (DT) vs. the non-drug target (NDT) driven nodes generated by our algorithms, the average (per node) control efficacy of the former ones is considerably higher.

Furthermore, we analyzed the topological properties of the driven DT proteins and of the essential proteins in all cancer networks, in order to understand the structural and functional properties of these proteins. We observed that driven DT associated nodes have high degree in the network, Fig. 3, which shows that these proteins are central, and that they form robust connections inside the networks. This characteristic seems to confirm the control efficiency of these nodes, as it shows that these proteins have multiple connections within the networks and this intensifies the feasible control over the target (essential) nodes. Also, we observed that the essential proteins have high betweenness centrality, Fig. 3, showing that these proteins operate as a bridge in the networks, and that they are highly important for the signal flow.

We analyzed the relationship between our resulted driven DT proteins and cancer. We observed that some of these DT proteins are oncoproteins, and thus the associated targeting drugs have a strong potential therapeutic effect in those cancers, Table 3. Other driven DT proteins based drugs are known to be part of therapies in other cancers, but not in breast, pancreatic or ovarian. We also observed that out of all selected 75 driven DT proteins in all three cancer networks, 31 DT proteins are present in more than one cancer. This shows that some DT proteins are expressed in multiple cancer cell lines. We also analyzed the functional properties of high control proteins in all cancers, and found that these proteins are directly responsible for the occurrence of particular cancers.

The control methodology applied in this study provides an efficient way to control an interactome network through known drug target nodes, especially in the case of disease associated networks. Also, this work provides a better understanding of the disease associated biochemical networks and opens a new way towards the successful application of drug-target based control mechanisms. This in turn could pave the way for future studies of various disease diagnostic techniques based on network controllability, efficient therapeutic approaches, and personalized medicine.

Supporting Information

S1 Fig

Breast cancer PPI network. A yeast two-hybrid (Y2H) PPI network for cancer signalling proteins. A network view of pancreatic cancer PPI network. This network containd 90% of total network nodes, the remaining part of the network containing isolated nodes. The drug-target nodes (DTN) are shown in dark blue, target nodes (TN) are in maroon, nodes that are both in DTN and TN are shown in dark green, Non-drug target nodes (NDT) are in light orange, and nodes that both in NDT and TN are in magenta.

S2 Fig

Ovarian cancer PPI network. A yeast two-hybrid (Y2H) PPI network for cancer signalling proteins. A network view of pancreatic cancer PPI network. This network containd 90% of total network nodes, the remaining part of the network containing isolated nodes. The drug-target nodes (DTN) are shown in dark blue, target nodes (TN) are in maroon, nodes that are both in DTN and TN are shown in dark green, Non-drug target nodes (NDT) are in light orange, and nodes that both in NDT and TN are in magenta.

S1 Table

Breast, Pancreatic and Ovarian cancer proteins.

S2 Table

PPI network of breast, pancreatic and ovarian cancer.

S3 Table

Essential proteins of breast, pancreatic and ovarian cancer.

S4 Table

Drug-target proteins.

Acknowledgments

This work was supported by the Academy of Finland through grant 272451.

References

[1] Sever R, Brugge JS. Signal Transduction in Cancer. Cold Spring Harb Perspect Med. 2015; 5:a006098. doi:10.1101/cshperspect.a006098

[2] Pawson T, NAsh P. Protein-protein interactions define specific in signal transduction. Genes & Dev. 2000 May 1; 14(9):1027-1047. doi:10.1101/gad.14.9.1027

[3] Warna JL. Signaling by the TGFβ superfamily. Cold Spring Harb Perspect Biol. 2013; 5:a011197. doi:10.1101/cshperspect.a011197

[4] Mu Y, Gudey SK, Landstrm M. Non-Smad signaling pathways. Cell Tissue Res. 2012 Jan; 347(1):11-20. doi:10.1007/s00441-011-1201-y

[5] Elston R, Inman GJ. Crosstalk betweeen p53 and TGF-beta signalling. J Signal Transduct. 2012; 2012:294097. doi:10.1155/2012/294097

[6] Pawson T, Warner N. Oncogenic re-wiring of cellular signaling pathways. Oncogene. 2007; 26:1268-1275. doi:10.1038/sj.onc.1210255

[7] Amit I, Wides R, Yarden Y. Evolvable signaling networks of receptor tyrosine kinases: relevance of robustness to malignancy and to cancer therapy. Molecular Systems Biology. 2007; 3:151. doi:10.1038/msb4100195

[8] Alexander XM, Huang PH. Receptor Tyrosine Kinase Coactivation Networks in Cancer. Cancer Res. 2010 May 15; 70(10):3857-60. doi:10.1158/0008-5472.CAN-10-0163

[9] Kolch W, Halasz M, Granovskaya M, Kholodenko NB. The dynamic control of signal transduction networks in cancer cells. Nat Rev Cancer. 2015 Sep; 15(9):515-27. doi:10.1038/nrc3983.

[10] Menche J, Sharma A, Kitsak M, Ghiassian DS, Vidal M, Loscalzo J, Barabsi LA. Uncovering disease-disease relationships through the incomplete interactome. Science. 2015 Feb 20; 347:1257601; doi:10.1126/science.1257601

[11] Shim EJ, Lee I. Network-assisted approaches for human disease research. Animal Cells and Systems. 2015; 4:231-235; doi:10.1080/19768354.2015.1074108

[12] Goh K-II, Choi IG. Exploring the human diseasome: the human disease network. BRIEFINGS IN FUNCTIONAL GENOMICS. 2012; 6:533-542. doi:10.1093/bfgp/els032

[13] Goh K-II, Cusick EM, Valle D, Childs B, Vidal M, Barabsi LA. The human disease network. PNAS. 2007 May 22; (104)21:8685-8690. doi:10.1073/pnas.0701361104

[14] Erler JT, Linding R. Network-based drugs and biomarkers. J Pathol. 2010 Jan; 220(2):290-6. doi:10.1002/path.2646

[15] Farkas IJ, Korcsmros T, Kovcs IA, Mihalik A, Palotai R, Simk GI, Szalay KZ, Szalay-Bek M, Vellai T, Wang S, Csermely P. Network-based tools for the identifi cation of novel drug targets. Sci. Signal. 2011 May 11; 4:173. doi:10.1126/scisignal.2001950

[16] Korcsmros T, Szalay MS, Bde C, Kovcs IA, Csermely P. How to design multi-target drugs: Target-search options in cellular networks. Expert Opin. Drug Discov. 2007; 2:110. doi:10.1517/17460441.2.6.799

[17] Csermely P, goston V, Pongor S. The efficiency of multi-target drugs: The network approach might help drug design. Trends Pharmacol. Sci. 2005; 26:178182. doi:http://dx.doi.org/10.1016/j.tips.2005.02.007

[18] Clatworthy AE, Pierson E, Hung DT. Targeting virulence: a new paradigm for antimicrobial therapy. Nat Chem Biol. 2007 Aug 20; 3:5418. doi:10.1038/nchembio.2007.24

[19] Zhan T, Boutros M. Towards a compendium of essential genes From model organisms to synthetic lethality in cancer cells. Crit Rev Biochem Mol Biol. 2015 Dec 01; 51(2): 7485. doi:10.3109/10409238.2015.1117053

[20] Wang T, Birsoy K, Hughes WN. Identification and characterization of essential genes in the human genome. Science. 2015 Nov; 350(6264): 10961101. doi:10.1126/science.aac7041

[21] Pyatnitskiy M, Karpov D, Poverennaya E, Lisitsa A, Moshkovskii S. Bringing Down Cancer Aircraft: Searching for Essential Hypomutated Proteins in Skin Melanoma. PLoS ONE. 2015; 10(11):e0142819. doi:10.1371/journal.pone.0142819

[22] Gao J, Liu Y, D'Souza MR, Barabsi LA. Target control of complex networks. Nature Communications. 2014 Nov 12; 5:5415. doi:10.1038/ncomms6415

[23] Kolch W, Halasz M, Granovskaya M, Kholodenko NB. The dynamic control of signal transduction networks in cancer cells. Nat Rev Cancer. 2015 sep; 15(9):515-27. doi:10.1038/nrc3983

[24] Zaudo JGT, Albert R. Cell Fate Reprogramming by Control of Intracellular Network Dynamics. PLoS Comput Biol. 2015; 11(4): e1004193. doi:10.1371/journal.pcbi.1004193

[25] Liu Y, Slotine j, Barabsi LA. Controllability of complex networks. Nature. 2011 May 12; 473:167-73. doi:10.1038/nature10011

[26] Wuchty S. Controllability in protein interaction networks. Proceedings of the National Academy of Sciences. 2014 May; 111(19):7156-60. doi:10.1073/pnas.1311231111

[27] Khuri S, Wuchty S. Essentiality and centrality in protein interaction networks revisited. BMC Bioinformatics. 2015; 16:109. doi:10.1186/s12859-015-0536-x

[28] UniProt Consortium. UniProt: a hub for protein information. Nucleic Acids Res. 2015 Jan; 43(Database issue):D204-12. doi:10.1093/nar/gku989

[29] Adjo A J, Lin S-X. Comparison of Functional Proteomic Analyses of Human Breast Cancer Cell Lines T47D and MCF7. PLoS ONE. 2012; 7(2): e31532. doi:10.1371/journal.pone.0031532

[30] Lee M-K, Nam K, Oh S. Extracellular matrix protein 1 regulates cell proliferation and trastuzumab resistance through activation of epidermal growth factor signaling. Breast Cancer Research; 2014 Dec 11; 16:479. doi:10.1186/s13058-014-0479-6

[31] Britton D, Zen Y, Quaglia A, Selzer S, Mitra V, et al. Quantification of Pancreatic Cancer Proteome and Phosphorylome: Indicates Molecular Events Likely Contributing to Cancer and Activity of Drug Targets. PLoS ONE. 2014; 9(3): e90948. doi:10.1371/journal.pone.0090948

[32] Haun RS, Fan CY, Mackintosh SG, Zhao H, Tackett AJ. CD109 Overexpression in Pancreatic Cancer Identified by Cell-Surface Glycoprotein Capture. J Proteomics Bioinform. 2014; S10:003. doi:10.4172/jpb.S10-003

[33] Hofmann TB, Schluter L, Lange P. COSMC knockdown mediated aberrant O-glycosylation promotes oncogenic properties in pancreatic cancer. Molecular Cancer. 2015; 14:109. doi:10.1186/s12943-015-0386-1

[34] Shahinian H, Loessner D, Biniossek LM. Secretome and degradome profiling shows that Kallikrein- related peptidases 4, 5, 6, and 7 induce TGFb-1 signaling in ovarian cancer cells. Molecular Oncology. 2014 Feb 1; 8:68-82. doi:10.1016/j.molonc.2013.09.003

[35] The Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. Nature. 2011 June 30; 474:609-15. doi:10.1038/nature10166

[36] Domcke S, Sinha R, Levine AD. Evaluating cell lines as tumour models by comparison of genomic profiles. NATURE COMMUNICATION. 2013 July 09; 4:2126. doi: 10.1038/ncomms3126

[37] Perfetto L, Briganti L, Calderone A. SIGNOR: a database of causal relationships between biological entities. Nucl. Acids Res. 2016 Jan 4; 44(D1):D548-54. doi:10.1093/nar/gkv1048

[38] http://combio.abo.fi/research/network-controlability-project/.

[39] Wang T et al. Identification and characterization of essential genes in the human genome. Science 2015; 350(6264):1096-1101. doi:10.1126/science.aac7041

[40] Marcotte R, Brown RK, Suarez F et al. Essential Gene Profiles in Breast, Pancreatic, and Ovarian Cancer Cells. Cancer Discovery. 2012 Feb; 2(2):172-89. doi:10.1158/2159-8290

[41] Zhan T, Boutros M. Towards a compendium of essential genes - From model organisms to synthetic lethality in cancer cells. Crit. Rev. Biochem. Mol. Biol. 2016; 51(2):74-85 doi:10.3109/10409238.2015.1117053

[42] Koh YLJ,Brown RK,Sayad A,Kasimer D,Ketela T,Moffat1 J. COLT-Cancer: functional genetic screening resource for essential genes in human cancer cell lines. Nucl. Acids Res. 2012 Jan; 40(Database issue): D957D963. doi:10.1093/nar/gkr959

[43] Law V, Knox C, Djoumbou Y, Jewison T. DrugBank 4.0: shedding new light on drug metabolism. Nucl. Acids Res. 2014; 1-7 doi:10.1093/nar/gkt1068

[44] Lin, C.-T. Structural controllability. IEEE Trans. Automat. Contr. 1974; 19:201-208; doi:10.1109/TAC.1974.1100557

[45] Blackhall L, Hill DJ. On the structural controllability of networks of linear systems (2nd IFAC Workshop on Distributed Estimation and Control in Networked Systems). 2010; 245-250.doi:10.3182/20100913-2-FR-4014.00079

[46] Murota, K., Poljak, S. Note on a graph-theoretic criterion for structural output controllability. IEEE T. Automat. Contr. 1990; 35:939-942; doi:10.1109/9.58507

[47] Yu L. High-quality binary protein interaction map of the yeast interactome network. Science. 2008 Oct 3; 3:322(5898):104-10. doi:10.1126/science.1158684

[48] He X, Zhang J. Why Do Hubs Tend to Be Essential in Protein Networks? PLoS Genet. 2006 June 2; 2(6):e88. doi:10.1371/journal.pgen.0020088

[49] Barabsi AL, Oltavi ZN. Network biology: understanding the cells functional organization. Nat Rev Genet. 2004 Feb; 5(2):101-13. doi:10.1038/nrg1272

[50] Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics. PLoS Comput Biol. 2011; 3(4):e59. doi:10.1371/journal.pcbi.0030059

[51] Badache, A, Gonalves, A. The ErbB2 signaling network as a target for breast cancer therapy. J Mammary Gland Biol Neoplasia. 2006 Aug. 11(1):15-25. doi: 10.1007/s10911-006-9009-1

[52] Gao S, Chen X, Jin H. Overexpression of ErbB2 renders breast cancer cells susceptible to 3-BrPA through the increased dissociation of hexokinase II from mitochondrial outer membrane. Oncology Letters. 2016 Dec 21; 11:1567-73. doi:10.3892/ol.2015.4043

[53] Balko, J., I. Mayer, M. Levy, C. Arteaga. 2013. HER2 (ERBB2) in Breast Cancer. My Cancer Genome http://www.mycancergenome.org/content/disease/breast-cancer/erbb2/ (Updated April 10).

[54] Carmona JF, Montemurro F, Kannan S, Rossi V, Verma C, Baselga J, Scaltriti M. AKT signaling in ERBB2-amplified breast cancer. Pharmacol Ther. 2016 Feb; 158:63-70. doi:10.1016/j.pharmthera.2015.11.013

[55] Finn SR. Targeting Src in breast cancer. Annals of Oncology. 2008 May 16; 19(8):13791386. doi:10.1093/annonc/mdn291

[56] Sen B, Johnson MF. Regulation of Src Family Kinases in Human Cancers. J Signal Transduct. 2011; 2011:865819, doi:10.1155/2011/865819

[57] Fyffe C. Falasca M. 3-Phosphoinositide-dependent protein kinase-1 as an emerging target in the management of breast cancer. Cancer Manag Res. 2013 Aug 23; 5:271280. doi:10.2147/CMAR.S35026

[58] Maurer M, Su T, Saal LH, Koujak S, Hopkins BD, Barkley CR, Wu J, Nandula S, Dutta B, Xie Y, Chin YR, Kim DI, Ferris JS, Gruvberger-Saal SK, Laakso M, Wang X, Memeo L, Rojtman A, Matos T, Yu JS, Cordon-Cardo C, Isola J, Terry MB, Toker A, Mills GB, Zhao JJ, Murty W, Hibshoosh H, Parson R. 3-Phosphoinositide-dependent kinase 1 potentiates upstream lesions on the phosphatidylinositol 3-kinase pathway in breast carcinoma. Cancer Res. 2009 Aug 1; 69:62996306. doi:10.1158/0008-5472.CAN-09-0820

[59] Cimino D. Identification of new genes associated with breast cancer progression by gene expression analysis of predefined sets of neoplastic tissues. Int J Cancer. 2008 Sep 15; 123(6):1327-38. doi:10.1002/ijc.23660.

[60] Wheler JJ, Atkins JT, Janku F, Moulder SL, Yelensky R, Stephens PJ, Kurzrock R. Multiple gene aberrations and breast cancer: lessons from super-responders. BMC Cancer. 2015 May 29; 15:442. doi:10.1186/s12885-015-1439-y.

[61] Zhou Z, Patel M. Identification of synthetic lethality of PRKDC in MYC-dependent human cancers by pooled shRNA screening. BMC Cancer. 2014 Dec 14; 14:944. doi:10.1186/1471-2407-14-944

[62] Paplomata E, ORegan R. The PI3K/AKT/mTOR pathway in breast cancer: targets, trials and biomarkers. Ther Adv Med Oncol. 2014 Jul; 6(4):154166. doi:10.1177/1758834014530023

[63] McAuliffea FP, Meric-Bernstama F, Millsb BG, Gonzalez-Angulob MA. Deciphering the Role of PI3K/Akt/mTOR Pathway in Breast Cancer Biology and Pathogenesis. Clinical Breast Cancer. 2010; 10(3):S59-65. doi:10.3816/CBC.2010.s.013

[64] Kolb A, Kleeff J, Arnold N. Expression and differential signaling of heregulins in pancreatic cancer cells. Int. J. Cancer. 2007 Feb 1; 1:120(3):514-23. doi:10.1002/ijc.22360

[65] Skrypek N, Vasseur R. Vincent A. The oncogenic receptor ErbB2 modulates gemcitabine and irinotecan/SN-38 chemoresistance of human pancreatic cancer cells via hCNT1 transporter and multidrug-resistance associated protein MRP-2. Oncotarget. 2015 May 10; 6(13):1085310867. doi:10.18632/oncotarget.3414

[66] Kelber AJ, Reno T, Kaushal S. KRas Induces a Src/PEAK1/ErbB2 Kinase Amplification Loop That Drives Metastatic Growth and Therapy Resistance in Pancreatic Cancer. Cancer Res. 2012 May 15; 72(10):25542564. doi:10.1158/0008-5472.CAN-11-3552

[67] Albury MT, Pandey V, Gitto BS. Constitutively Active Akt1 Co-operates with KRasG12D to Accelerate In Vivo Pancreatic Tumor Onset and Progression. Neoplasia. 2015 Feb; 17(2):175182. doi:http://dx.doi.org/10.1016/j.neo.2014.12.006

[68] Shi P, Yin T, Zhou F. Valproic acid sensitizes pancreatic cancer cells to natural killer cell-mediated lysis by upregulating MICA and MICB via the PI3K/Akt signaling pathway. BMC Cancer. 2014 May 14; 14:370. doi:10.1186/1471-2407-14-370

[69] Marchand B, Arsenault D, Raymond-Fleury A, Boisvert FM, Boucher MJ. Glycogen synthase kinase-3 (GSK3) inhibition induces prosurvival autophagic signals in human pancreatic cancer cells. J Biol Chem. 2015 Feb 27; 290(9):5592-605. doi:10.1074/jbc.M114.616714

[70] Jacobs MK, Bhave RS, Ferraro RS. GSK-3: A Bifunctional Role in Cell Death Pathways. International Journal of Cell Biology. 2012; 2012:930710, doi.org/10.1155/2012/930710

[71] Kim PG, Billadeau DD. GSK-3β Inhibition in Pancreatic Cancer. Pancreatic Cancer. 2008; 635-46. doi:10.1007/978-0-387-69252-4-37

[72] Angelescu R, Gheonea ID, Nitulescu C. ABL1 mRNA levels in pancreatic cancer and chronic pancreatic. Annals of RSCB. 2012. Vol. XVII, 1

[73] Greuber KE, Smith-Pearson P, Wang J. Role of ABL Family Kinases in Cancer: from Leukemia to Solid Tumors. Nat Rev Cancer. 2013; 13(8): 559571. doi:10.1038/nrc3563

[74] Levav-Cohen Y, et al. C-Abl as a modulator of p53. Biochem. Biophys. Res. Commun. 2005 Jun 10; 331:73749. doi:10.1016/j.bbrc.2005.03.152

[75] Ishizuka Y, Shimura M, Ishizuka Y. Expression of the Wild Type Rearranged during Transfection Protooncogene in Ovarian Cancer. Jikeikai Med J. 2011; 58:57-62

[76] Horio M, Kato T, Mii S. Expression of RET finger protein predicts chemoresistance in epithelial ovarian cancer. Cancer Medicine. 2012 Oct 1; 1(2):218-29. doi:10.1002/cam4.32

[77] Ma Y, Wei Z. Downregulation of TRIM27 expression inhibits the proliferation of ovarian cancer cells in vitro and in vivo. Laboratory Investigation. 2016 Jan; 96:3748; doi:10.1038/labinvest.2015.132

Table 3: **Highly impact drug-target proteins for Breast, Pancreatic and Ovarian cancers**

Cancer Types	Drug-target	Target proteins	Anti-cancer drug	Known to be used in cancer therapies
Breast	ERBB2	CDK1, CDCH2, CDC7, SH3RF1, APLP2	Lapatinib	Breast, Lung
	SRC	PLK1, RAN, MAP2K1, KARS	Dasatinib, Bosutinib, Ponatinib	Chronic myelogenous leukemia (CML)
	PDPK1	PNK1, ERBB3, SH3RF1, PDPK1	None	None
	PRKDC	GBF1, MN1, RPA2	None	None
	MTOR	PHB2, RPTOR, MTOR	Temsirolimus	Renal cell carcinoma (RCC), Bone marrow cancer,
	JAK2	MAP3K5, AIRE	Ruxolitinib, Erlotinib	Pancreatic cancer and others types of cancer
	HDAC3	SP1, HDAC3	Vorinostat	Cutaneous T cell lymphoma (CTCL)
	CDK2	PFN1, TFCP2	None	None
Pancreatic	ERBB2	TUBA1C, ERF, NUDC, ERBB2	Lapatinib	Breast, Lung
	AKT1	CNSK1E, MST1R, MAML1, ADAM17	None	None
	GSK3B	DLC1, ROBO1, ABL1	Hepatitis B immune globulin, Alectinib, Paclitaxel, Eribulin, Testolactone	Liver cancer, Anaplastic lymphoma kinase (ALK) and Non-small cell lung cancer (NSCLC), Cancer chemotherapy, Breast cancer
	ABL1	DLC1, ROBO1, ABL1	None	None
	IGF1R	PIK3C2A, IGFR1	None	None
	HDAC3	SMURF2, HDAC3	Vorinostat	Cutaneous T cell lymphoma (CTCL)
	RAF1	STK3, DAXX	None	None
	INSR	IRS4, INSR	None	None
	RAC1	SFN, USP6	None	None
	PDPK1	HNRNPA1, PDPK1	None	None
Ovarian	RET	MAPK3, PLK1, OPTN, PTTG1, CDH1	Cabozantinib, Lenvatinib, Vandetanib, Sunitinib, Regorafenib, Ponatinib,	Medullary thyroid cancer (MTC), Thyroid cancer, Renal cell carcinoma (RCC), Imatinib-resistant gastrointestinal stromal tumor (GIST), Metastatic colorectal cancer and Advanced gastrointestinal stromal tumours, Chronic myeloid leukemia,
	AKT1	WNK1, CHEK1	None	None
	GRB2	APBB1, GRB2	None	None
	JAK3	STAT2, JAK3	None	None
	PRKDC	HNRNPU, VHL	None	None
	SMO	GNG12, GNAT2	Vismodegib, Sonidegib	Basal cell carcinoma
	MTOR	ISCU, RPS6	Temsirolimus	Renal cell carcinoma (RCC)
	CDK2	MYBL2, CHEK1	None	None

The columns represent the type of cancer, drug-target, target (essential) proteins, name of anti-cancer drug, and type of cancer for which the drug is known to be used.

TURKU

CENTRE *for*

COMPUTER

SCIENCE

Joukahaisenkatu 3-5 A, 20520 TURKU, Finland | www.tucs.fi

University of Turku

Faculty of Mathematics and Natural Sciences

- Department of Information Technology
- Department of Mathematics and Statistics

Turku School of Economics

- Institute of Information Systems Sciences

Åbo Akademi University

- Computer Science
- Computer Engineering

ISBN 978-952-12-3373-9
ISSN 1239-1891