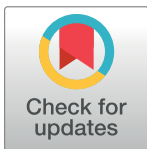RESEARCH ARTICLE

Hierarchical closeness-based properties reveal cancer survivability and biomarker genes in molecular signaling networks

Tien-Dzung Tran[1]¤*, Yung-Keun Kwon[2]*

1 Complex Systems and Bioinformatics Lab, Hanoi University of Industry, Hanoi, Viet Nam, **2** School of IT Convergence, University of Ulsan, Ulsan, Republic of Korea

¤ Current address: Department of Software Engineering, Faculty of Information and Communication Technology, Hanoi University of Industry, Hanoi, Viet Nam
* trantd.vn@gmail.com (TDT); kwonyk@ulsan.ac.kr (YKK)

Abstract

Specific molecular signaling networks underlie different cancer types and quantitative analyses on those cancer networks can provide useful information about cancer treatments. Their structural metrics can reveal survivability of cancer patients and be used to identify biomarker genes for early cancer detection. In this study, we devised a novel structural metric called hierarchical closeness (HC) entropy and found that it was negatively correlated with 5-year survival rates. We also made an interesting observation that a network of higher HC entropy was likely to be more robust against mutations. This finding suggested that cancers of high HC entropy tend to be incurable because their signaling networks are robust to perturbations caused by treatment. We also proposed a novel core identification method based on the reachability factor in the HC measure. The cores were permitted to decompose such that the negative relationship between HC entropy and cancer survival rate was consistently conserved in every core level. Interestingly, we observed that many promising biomarker genes for early cancer detection reside in the innermost core of a signaling network. Taken together, the proposed analyses of the hierarchical structure of cancer signaling networks may be useful in developing future novel cancer treatments.

Introduction

Cancer is a leading cause of disease worldwide with more than 11 million people diagnosed every year. It is estimated that by 2030 there will be approximately 26 million new cancer cases and 17 million cancer deaths worldwide per year [1]. Cancer is a genetic disease where one or more mutated genes result in abnormal cell proliferation. Early diagnosis and personalized therapy often rely on insights from relevant molecular signaling pathways as well as cancer-related genes. In terms of network dynamics, a signaling network converges on a stable equilibrium state (an ordinary attractor), which corresponds to a normal cellular state, but a genetic mutation may attract a cell to a malignant phenotypic state (a cancer attractor), which eventually results in cancer development [2]. Although a perturbation of a gene or an

interaction can be treated to recover the ordinary attractor, the cancer network may be strongly robust to the perturbation. Intriguingly, it has been reported that network robustness is highly related to a variety of structural characteristics, including feedback loops, auto-regulation, feed-forward loops, source—sink gradients, modularity, redundancy, and parallel pathways [3–7]. Therefore, the degree to which the malignant status is maintained against therapeutic perturbations can be related to various structural characteristics in cancer signaling networks. Previous studies have explored this idea [8–10]. In one study, the 5-year survival rates of patients diagnosed with one of 14 different types of cancers were shown to be negatively correlated with the degree of heterogeneous connectivity in all cancer signaling networks except the signaling network underlying prostate cancer [8]. A follow-up study proposed network modularity as a factor in patient survivability and indeed showed that the degree of modularity was negatively correlated with survivability in the same 14 cancers, including prostate cancer [9]. However, network modularity is not an easy-to-compute metric because a non-deterministic algorithm should be used to optimize the metric [11–13]. Taken together, it is more suitable to create a novel metric that can more accurately predict cancer survivability with little computational cost. Thus, we propose a novel metric based on a hierarchical structure.

Our new metric also relates to the core-periphery structural identification of directed networks. It is known that a number of molecular biological networks exhibit a core-periphery structure, wherein a few nodes are highly interconnected and the remaining nodes are loosely connected. Core nodes are identified by pruning underutilized links until core-peripheries emerge [14] and play a differing role from peripheral nodes [15, 16]. For example, in protein contact networks deleterious mutations which may cause cancer are more likely to be distributed near core nodes than peripheral nodes [17]. For instance, some prognostic genes of hepatocellular carcinoma, such as *B1* and *Sec62*, often reside at the core of gene co-expression networks [18]. Among the most recently proposed core identification methods [14, 19–21], *K-core* [22], which is based on the connectivity of a node, has been the most widely used method and was used to show that the core captures the most important characteristics of a network [23–25]. For example, biomarkers were found inside the core area of a co-expression network associated with liver cirrhosis [26] and a protein-protein interaction network associated with hepatocellular carcinoma [27]. However, *K-core* identification is not effective in the analysis of directed networks because the direction of interactions is not considered in the analysis. Therefore, a core identification method adequate for directed networks such as molecular signaling networks is required.

Hierarchical closeness (HC) is a heterogeneous network centrality measure proposed in our previous study [28]. We extend on this previous study in two ways. First, we suggest that HC values of entropy are a novel structural measure to evaluate cancer survivability, and we found that the HC entropy values were negatively correlated with patient 5-year cancer survival rates in 16 types of cancers. This relationship was more clear than in cases using degree entropy [8] and modularity [9]. We also found that HC entropy values were positively correlated with network robustness when extensive simulations based on random Boolean networks were utilized. Since network robustness represents how likely a network is to maintain an equilibrium state against perturbations, this result explains why cancers with relatively high HC entropy values tend to be incurable. Second, we exploit the reachability property to profile the core-periphery structure of directed networks and propose a novel network decomposition called *R-core*. Interestingly, the core area identified by *R-core* maintained the negative relationship of HC entropy to the cancer survival rate whereas the core area identified by the *K-core* decomposition did not. Furthermore, we discovered that many candidate cancer biomarker genes reside in the innermost core area identified by *R-core*, and most of those candidate

biomarker genes were validated empirically. This demonstrates that our method identified the functional central region of a directed molecular network.

Materials and methods

Cancer signaling networks and 5-year survival statistics

We used 16 pathways downloaded from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (www.genome.jp/kegg) by adding two additional pathways of the Gastro and the Breast cancers to those in the previous studies [8, 9]. KEGG includes comprehensive pathways manually derived from textbooks, literature, other databases, and expert knowledge and provides consensus information regarding the "cancer site" which refers to the tissue or cell type of the primary tumor. We did not consider other pathway databases such as BioCyc (www.biocyc.org), Reactome (www.reactome.org), and BioGRID (www.thebiogrid.org) for analysis since they do not include pathways corresponding to a specific cancer site.

Each cancer pathway is represented by a network, wherein a node and an edge correspond to a protein and an interaction between proteins, respectively. In the network, directed edges represent activation, inhibition, expression, indirect-effect, interaction via compound, missing-interaction, and phosphorylation whereas undirected edges represent protein-protein interactions such as binding/association and dissociation. All types of interactions were included in structrual analysis about HC entropy and K-core/R-core decomposition by interpreting undirected edges as bi-directional interactions. However, we note that only activation and inhibition types of interactions were considered for network robustness estimation as done by previous studies [29, 30]. It should be also noted that the original KGML (KEGG Markup Language) files downloaded from KEGG pathway database were not consistent to the static pathway map image, and we corrected this inconsistency by using the Cytoscape plug-in KEGGParser [31]. The patient 5-year survival rates for every cancer type were obtained from the Surveillance Epidemiology and End Results (SEER) Program database (www.seer.cancer.gov), which is a resource for epidemiological data compiled by the U.S. National Cancer Institute (www.cancer.gov).

Hierarchical-closeness (HC) entropy

There have been a number of different entropy measures proposed for general network analyses [32–36]. In systems biology, some network entropies were used to identify dynamic changes in time course differentiation data and to predict higher levels of cellular plasticity in cancer stem cell populations [37]. Among them, degree entropy was proposed to be a significant cancer system property in undirected networks such as in a protein-protein interaction network [38]. The Shannon entropy of a discrete random variable X, $H(X)$, is defined as

$$H(X) = -\sum_{x \in X} p(x) \log_2 p(x) \tag{1}$$

where $p(x)$ is the probability mass function of X. The degree entropy of a network is obtained by letting X represent degree values of the network, where each value $x \in X$ is the number of interactions incident to a node. In this study, we derived another network entropy, HC entropy, by letting X represent HC values of the network, where each value $x \in X$ is the HC value of a node. The HC value of a node v, $HC(v)$, was defined previously [28] by combining reachability and closeness as

$$HC(v) = R(v) + C(v) \tag{2}$$

where $R(v)$ represents the reachability value of node v, which is the number of nodes in the

network that can be reachable from node v, and $C(v)$ is the closeness centrality normalized into $[0, 1]$ [39]. $C(v)$ is defined as

$$C(v) = \frac{1}{|V| - 1} \sum_{w \in V \setminus \{v\}} \frac{1}{d(v, w)} \tag{3}$$

where $|V|$ is the number of nodes, and $d(v, w)$ is the distance of the shortest path, if any, from v to w; otherwise, $d(v, w)$ is specified as an infinite value. Considering that $R(v)$ is an integer and $C(v) \in [0, 1)$, the definition of $HC(v)$ implies that all nodes $v \in V$ are first grouped by $R(v)$ and the nodes of a same group are further grouped by $C(v)$. Based on (1), (2), and (3), we computed HC entropies of 16 cancer pathways.

K-core and R-core decomposition

In general, networks can be decomposed into a dense core and a loosely connected periphery by utilizing a network decomposition method. *K-core* decomposition based on the node degree is often used [40] to identify particular subsets of a network, called *k-cores* ($k \geq 1$), where k denotes a core level. A *k-core* of a network G is composed of a subset of nodes in a network G, which is obtained by the following pruning rule. Given a network, all the nodes with a degree $< k$ are removed, along with their incident interactions, from the network. This removal process is repeated until the degree of every node in the remaining network is $\geq k$. The *k-core* denotes the remaining set of nodes, and hence it holds that k_1-core is a subset of k_2-core if $k_1 \geq k_2$. In this study, we suggest another network decomposition, *R-core*, which is based on the reachability value $R(v)$. It employs the same pruning rule as the *K-core* decomposition method except $R(v)$ is used instead of the node degree. In other words, all nodes with $R(v) < r$ and their incident interactions are removed at every pruning step. As a result, *R-core* decomposes a directed network into sub-networks called *r-cores*. By the decomposition definition, *k-* and *r-cores* represent a greater inner core as the core level value increases. Furthermore, a *k-shell* (or *r-shell*) is defined as a set of nodes that belong to the *k-core* (or *r-core*) but not to $(k + 1)$-core (or $(r + 1)$-core). An example of these network decompositions is provided (Fig 1). For convenience, the *periphery* and the *outermost core* denote the shell and the core with the lowest level, respectively. In addition, *1-* and *2-innermost core* denote cores with the first and the second highest level, respectively.
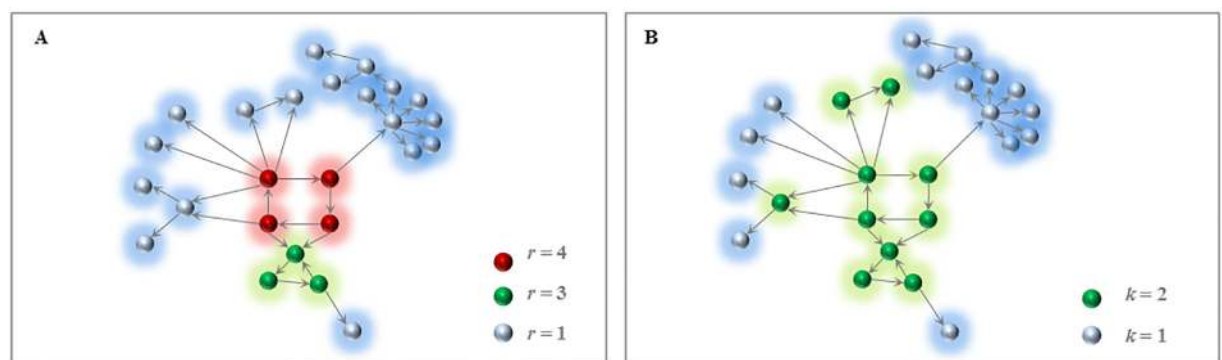
Fig 1. An example of *R-core* and *K-core* decompositions. A directed network with 25 nodes and 29 interactions is given. **(A)** Result of the *R-core* decomposition. Three *r-shells* ($r = 1$, 3, and 4; light blue, green, and red, respectively) were identified. **(B)** Result of the *K-core* decomposition. Two *k-shells* ($k = 1$ and 2; light blue and green, respectively) were identified.

https://doi.org/10.1371/journal.pone.0199109.g001

Network robustness

To evaluate network robustness of both cancer pathways and random networks, we employed a Boolean network model, which has been extensively used to investigate the dynamics of various signaling networks [41–44]. A signaling network is represented by a directed graph $G(V, A)$ where $V = \{v_1, v_2, \ldots, v_n\}$is a set of Boolean variables and A is a set of the directed interactions. Each v_i has a value of 1 ('on') or 0('off') which represents activated or inactivated statuses, respectively, and the state of v_i should be updated by a corresponding logical function f_i. In this study, either a logical conjunction or a disjunction is randomly selected for each f_i uniformly at random. For example, if a Boolean variable v has activation relationships with v_1 and v_2, and an inhibitory relationship with v_3, then the respective conjunction and disjunction update rules are $v(t + 1) = \wedge(v_1(t), v_2(t), \bar{v}_3(t))$ and $v(t + 1) = \vee(v_1(t), v_2(t), \bar{v}_3(t))$. In the case of the conjunction, the value of v at time $t + 1$ is 1 only if the values of v_1, v_2, and v_3 at time t are 1, 1, and 0, respectively. On the contrary, in the case of the disjunction, the value of v at time $t + 1$ is 1 if either $v_1(t) = 1$, $v_2(t) = 1$, or $v_3(t) = 0$ holds. In addition, the states of all genes are synchronously updated. Then a network state defined as $s(t) = (v_1(t), v_2(t), \ldots, v_n(t))$ at time t transits to the next state $s(t + 1)$ according to a set of update rules defined as $F = \{f_1, f_2, \ldots, f_n\}$(i.e., $s(t + 1) = F(s(t))$. The network eventually converges to a fixed state, or a limit-cycle attractor. We denote the attractor starting from state $s(t)$ as $\langle s(t) \rangle$. The network is called robust to a perturbation at v if $\langle s \rangle$ equals to $\langle s\bar{v} \rangle$ where $\bar{v}(= \neg v)$ indicates the state perturbation of s subject to v. This robustness concept has been widely used [45–47]. Specifically, the robustness of a network $\gamma(G)$ is defined as follows:

$$\gamma(G) = \frac{1}{n \cdot |S|} \sum_{v \in \mathbf{V}} \sum_{s \in \mathbf{S}} I(\langle s \rangle = \langle s\bar{v} \rangle) \tag{4}$$

where S is a set of whole network states (here, $|S| = 2^n$), and $I(\cdot)$ is an indicator function, it outputs 1 if $I(true)$ or 0 otherwise.

Results

HC entropy indicates cancer 5-year survival rates

A previous study found that the degree entropy of cancer signaling networks correlates to cancer patient survivability in all cases analyzed with the exception of prostate cancer [8]. Another previous finding showed that hierarchical closeness can effectively predict genes more susceptible to mutations [28]. These findings led us to investigate the correlation between patient 5-year survival rates and two entropy measures, degree entropy and HC entropy, of 16 cancer signaling networks (Fig 2). We observed a non-significant negative correlation between the survival rate and degree entropy (Fig 2A; Spearman's rank correlation coefficient $R_s = -0.294$, $P = 0.255$), which corroborates previous data [8]. However, HC entropy showed a significant negative relationship (Fig 2B; Spearman's rank correlation coefficient $R_s = -0.591$, $P = 0.022$). This comparison suggests that HC entropy is a more accurate measure of 5-year cancer survival rates. Topologies of two example cancer signaling pathways are also shown (Fig 3). Pancreatic cancer has a relatively high HC entropy (4.40) and a relatively low 5-year survival rate (0.055), while basal cell carcinoma has a relatively low HC entropy (2.65) and a relatively high 5-year survival rate (0.914). The nodes of pancreatic cancer signaling pathway tend to be more dispersed than the nodes of the basal cell carcinoma signaling pathway, which illustrates different HC entropies.
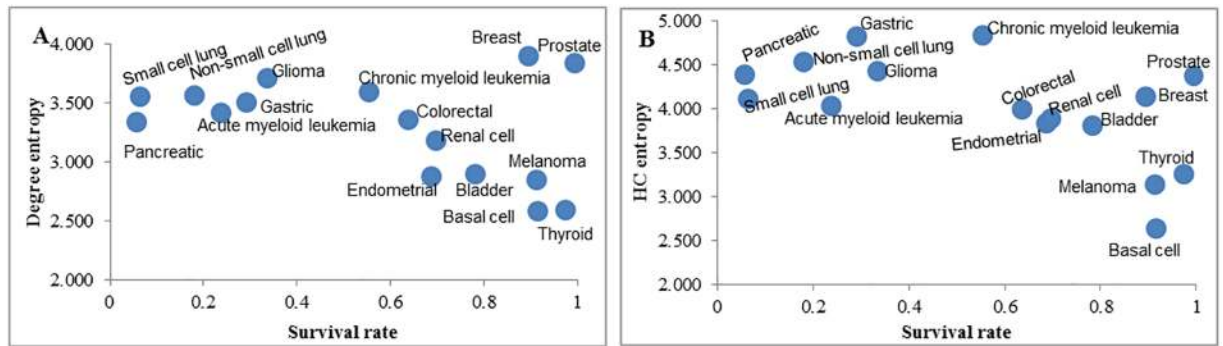
Fig 2. The correlation of degree entropy and HC entropy, respectively, with patient 5-year survival rates for 16 cancer signaling networks. (**A**) Result of degree entropy. A significant correlation is not observed (Spearman's rank correlation coefficient $R_s = -0.294$, $P = 0.255$). (**B**) Result of HC entropy. A significant negative correlation is indicated (Spearman's rank correlation coefficient $R_s = -0.591$, $P = 0.022$).

https://doi.org/10.1371/journal.pone.0199109.g002

HC entropy reflects robustness of cancer signaling networks

In the previous section, we showed that HC entropy of cancer signaling networks is negatively correlated with cancer survivability. Next, we assumed that HC entropy could be represented by how much a signaling network is robust to mutations and therefore built a Boolean network model to investigate the relationship between HC entropy and network robustness [7, 28]. To do this, we generated random networks using a previous shuffling method [48, 49], which iteratively chooses a pair of interactions (v_a, v_b) and (v_c, v_d) uniformly at random and replaces them with a pair of new interactions (v_a, v_d) and (v_c, v_b). We note that in-/out-degree distributions of the original network are preserved whereas the reachability distribution is not in the resultant random network. We constructed a pool of 80000 random networks by shuffling each of 16 cancer signaling pathways 5000 trials. For unbiased analysis in the distribution of robustness, we considered 24 equal-sized bins by the robustness values ranged from 0.52 to 1 by 0.02, and chose 20 networks out of the random network pool for every bin. As a result, we investigated the HC entropy and the network robustness values of 480 random networks. As
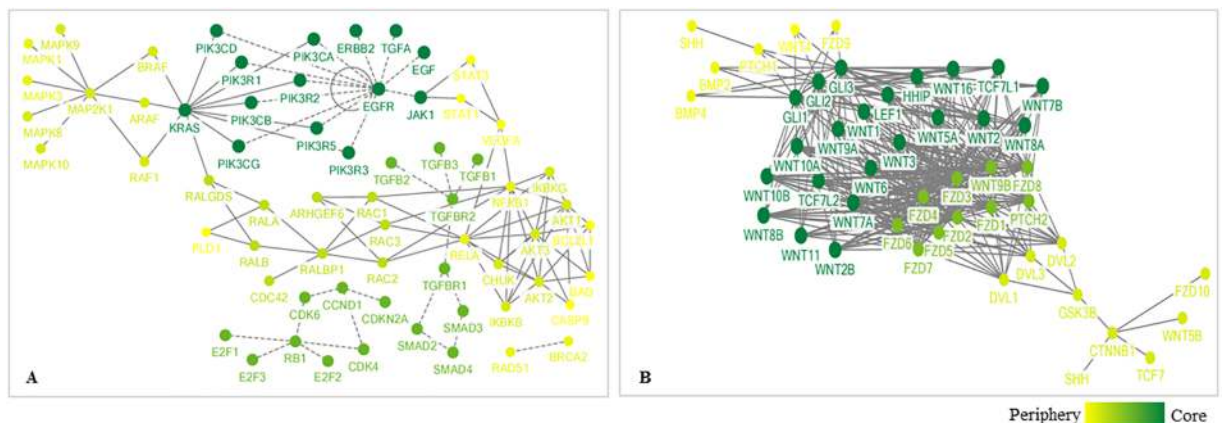
Fig 3. Topological visualization of two cancer signaling pathways with different HC entropy values. (**A**) A pancreatic cancer signaling network shows relatively high HC entropy (4.40) and a low patient 5-year survival rate (0.055). (**B**) A basal cell carcinoma signaling network shows relatively low HC entropy (2.65) and a high patient 5-year survival rate (0.914). The distribution of HC values is more heterogeneous in pancreatic cancer than in basal cell carcinoma networks (see S1 Fig). Dashed and solid lines represent undirected and directed interactions, respectively. The core-periphery structures are profiled by *R-core* decomposition. Dark green and light yellow circles represent nodes in shells of a higher and lower level, respectively. Node labels represent NCBI gene symbols.

https://doi.org/10.1371/journal.pone.0199109.g003
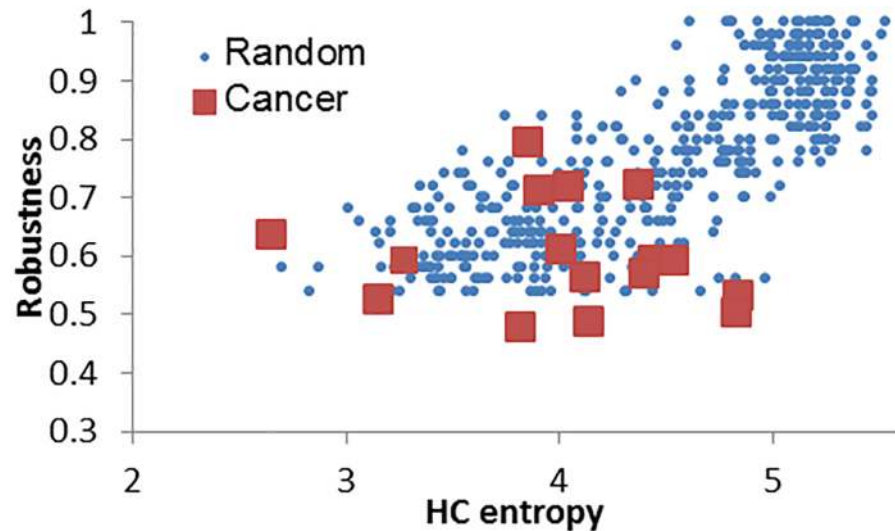
Fig 4. The positive correlation between HC entropy and network robustness in 16 cancer signaling pathways and in random networks. We first constructed a pool of 80000 random networks by shuffling each cancer signaling pathway 5000 times. We considered 24 equal-sized bins by the robustness values ranged from 0.52 to 1 by 0.02, and chose 20 random networks out of the pool for each bin. As a result, we examined HC and robustness values of 480 random networks. The Spearman's rank correlation coefficient between HC entropy and network robustness in the random networks was significantly positive (R_s = 0.799 with P < 0.0001). The squares represent the results of the 16 cancer signaling networks.

https://doi.org/10.1371/journal.pone.0199109.g004

shown in Fig 4, the robustness was positively correlated with the HC entropy in the random networks (Spearman's rank correlation coefficient R_s = 0.799, P <0.0001). In other words, we observed that the network robustness increased as the heterogeneity of HC values increased. This finding supports the hypothesis that a cancer with a high HC entropy value is likely to be incurable because the corresponding signaling network tends to be robust to therapeutic perturbations. This result is also related to previous observations that showed highly modularized biological networks are sensitive to perturbations because distributions of HC values in these networks are likely to be relatively uniform [7, 50]. Taken together, HC entropy is an interesting architectural characteristic of cancer signaling networks that can indicate network robustness as well as predict patient cancer survival rates.

The core of cancer signaling networks

It is important to profile the network core-periphery structure because the core can contain the properties of the whole system. To this point, we compared *K-core* and *R-core* decompositions by determining whether the negative relationship between HC entropy and the 5-year survivor rate was consistently observed within the identified cores. We divided the nodes in the 16 cancer networks into three core levels as determined by either the *K-core* or *R-core* decompositions: *outermost*, *2-innermost*, and *1-innermost* (see Materials and methods). Then we calculated the correlation coefficients between HC entropy and the 5-year survival rates in the three core levels (Fig 5). As depicted in the figure, the negative correlations between HC entropy and 5-year survival rates were preserved in all core levels identified by *R-core* decomposition. On the contrary, the *2-innermost* and *1-innermost* core levels identified by *K-core* decomposition showed positive correlations, which is significantly different from the negative correlation of whole network. In addition, we note that the average sizes (i.e., the average numbers of nodes) of *1-innermost* cores identified by *R-core* decomposition (Mean = 18.93; STDEV = 12.51) and *K-core* decomposition (Mean = 14.50; STDEV = 8.90) over the 16 cancer
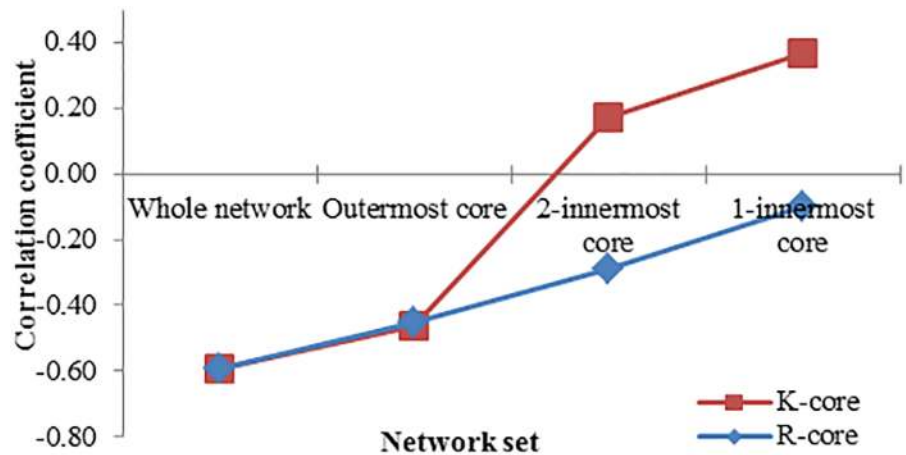
Fig 5. Comparisons of the negative correlation preservation (5-year survival rate, HC entropy) between *K-core* and *R-core* decompositions over 16 cancer signaling networks. Each cancer signaling network was decomposed into three core levels by *R-core* and *K-core* decompositions: *outermost*, *2-innermost*, and *1-innermost* (see Materials and methods for definitions). The correlation coefficient between the 5-year survival rate and the HC entropy of the nodes belonging to each core was respectively compared to that of whole network (−0.591). All the differences between the correlation coefficients of three cores by *R-core* decomposition and that of whole network were not significant (all P values > 0.05). On the other hand, the differences between the correlation coefficients of *2-innermost* and *1-innermost* cores by the *K-core* decomposition and that of whole network were significant (*P* = 0.03 and 0.007 in the case of *2-innermost* and *1-innermost* cores, respectively).

https://doi.org/10.1371/journal.pone.0199109.g005

networks were not significantly different from each other (P = 0.290). Taken together, the innermost core identified by *R-core* decomposition better captures the relationship of HC entropy to cancer survivability than that of *K-core* decomposition, despite the observation that both decompositions identified *1-innermost* cores of similar size.

Table 1. Candidate biomarker genes identified by *R-core* decomposition. In the table, **C1**, **C2**, and **C3** denote NCBI gene symbols of the three genes with the highest HC values among the *1-innermost* core genes as identified by *R-core* decomposition. The underlined genes mean they were previously reported as biomarker genes (see S1 Table).

Cancer site	Network properties				5-year survival rate	Candidate biomarker genes		
	The number of nodes	The number of edges	Degree entropy	HC entropy		C1	C2	C3
Acute myeloid leukemia	57	185	3.423	4.039	0.236	*FLT3*	*KIT*	*HRAS*
Basal cell carcinoma	47	534	2.588	2.646	0.914	*GLI1*	*GLI2*	*GLI3*
Bladder cancer	29	58	2.905	3.814	0.781	*RB1*	CDK4	*CCND1*
Breast cancer	144	773	3.905	4.138	0.897	*PGR*	NCOA3	NCOA1
Chronic myeloid leukemia	73	198	3.595	4.836	0.552	CRK	*CRKL*	*ABL1*
Colorectal cancer	49	153	3.365	4.002	0.636	*KRAS*	PIK3R5	*PIK3CA*
Endometrial cancer	44	121	2.884	3.845	0.686	EGF	*EGFR*	PIK3CA
Gastric cancer	148	682	3.506	4.824	0.306	*LRP5*	LRP6	WNT16
Glioma	64	255	3.719	4.437	0.334	*CCND1*	CDKN2A	*RB1*
Melanoma	68	719	2.853	3.148	0.912	*EGF*	FGF1	FGF2
Non-small-cell lung cancer	54	157	3.572	4.537	0.18	*KRAS*	*EGFR*	ERBB2
Pancreatic cancer	65	163	3.340	4.396	0.055	*KRAS*	*EGFR*	JAK1
Prostate cancer	86	404	3.836	4.371	0.994	*EGF*	IGF1	INS
Renal cell carcinoma	56	164	3.185	3.899	0.695	*HGF*	*MET*	GAB1
Small cell lung cancer	82	358	3.560	4.117	0.062	ITGB1	LAMC3	*COL4A1*
Thyroid cancer	28	73	2.601	3.267	0.972	*PPARG*	*PAX8*	RXRA

https://doi.org/10.1371/journal.pone.0199109.t001

Candidate cancer biomarker genes

Some recent studies reported that biomarker genes often reside in the innermost core of a biological network [18, 26, 27]. In addition, fragile genes, which are sensitive to replicative stress and exposure to environmental carcinogens, were also considered biomarkers in some cancers [51–53]. It has also been shown that a gene with a higher HC value is more likely to be a fragile gene in a human signaling network [28]. Guided by these results, we examined the three genes with the highest HC values in the *1-innermost* core as identified by *R-core* decomposition (Table 1). Interestingly, 33 genes out of a total of 48 genes were previously found to be biomarker genes for early cancer detection. For example, the three genes *KRAS*, *EGFR*, and *ERBB2*(*HER2*), which are found in the non-small cell lung cancer (NSCLC) signaling network, are key biomarkers [54, 55]. The *KRAS* gene is mutated in approximately 20% of NSCLC cases [56–58], and the *EGFR* gene is defective in approximately 10% of NSCLC patients and in nearly 50% of non-smoker lung cancer cases [59]. In addition, *HER2* mutations in NSCLC are present in approximately 4% of lung cancer patients [60]. Furthermore, the mutational frequency of the three genes may vary by race because of the impact of race/ethnicity on molecular pathways in human cancer [61, 62]]. For example, *EGFR* and *HER2* mutations among Korean lung cancer patients showed larger and lower frequencies, respectively, than that reported above [63]. Two additional genes, *CCND1* and *RB1*, identified in the bladder cancer signaling network are known biomarkers [64], and the third gene identified in the bladder cancer signaling network (*CDK4*) is a known inhibitor of bladder cancer [65, 66]. The top three genes found in the Glioma signaling network, *CCND1*, *CDKN2A*, and *RB1*, are frequently overexpressed, mutated, and/or deleted in glioma [67, 68]]. Molecular alterations of *EGFR* and *PIK3CA*, which are found in the endometrial cancer signaling network, have been reported in endometrial cancer studies [69, 70]. These examples suggest that the other top genes that have not been fully investigated yet may be promising candidate biomarkers.

Discussion

In this study, we investigated 16 types of cancer signaling networks and demonstrated that the hierarchical closeness entropy values and 5-year cancer survival rates were negatively correlated. This result suggested that when hierarchical closeness values of genes in a signaling network were more heterogeneous then these cancers are more likely incurable. SEER summaries of cancer survival rates do not account for cancer subtypes. Similarly, the KEGG pathways do not explicitly discern between cancer subtypes. For example, there are different subtypes in Melanoma cancer type, which have been characterized over the years: *BRAF* mutants [71, 72], *KRAS* mutants [73], *RAC1* mutants [74], *EGF* mutations [75], and other gene mutants. Patients affected by these subtypes have different survival times and respond differently to treatments. Therefore, we think that survival rate of a cancer patient is affected by two main factors: mutant position and signaling network structure. At a node level, mutations of a node at the innermost core are more deleteriously impactful than those at the rest of the network [17, 23, 76], thus such mutations may result in lower survival rates. At a network level, the network architecture with relatively high HC entropy results in lower cancer survival rates. The association between HC entropy and cancer survival rate at network structure level is a suggestion of prognostic model studies for cancer subtypes on the specific signaling pathways. Notably, the correlation between HC entropy and patient 5-year cancer survival rates was greater than the correlation between degree entropy and patient 5-year survival rates. In addition, HC entropy is a more attractive method than the modularity metric because of computational efficiency. Taken together, HC entropy can be considered an alternative to typical node-degree approaches for the associations between signaling network structure and cancer survivability.

To address the effects of HC entropy on patient 5-year survival rates, we simulated network robustness against mutations and found that the network robustness was positively correlated with the HC entropy of a network. This observation explains why cancer signaling networks with relatively high HC entropy are robust to therapeutic perturbations, which eventually make the cancer highly incurable. This result is also related to previous observations that showed highly modularized biological networks are sensitive to perturbations because distributions of HC values in these networks are likely to be relatively uniform [7, 50]. Two relationships (HC entropy, patient 5-year survival rate) and (HC entropy, network robustness) can be investigated further by a study of the correlation between patient 5-year survival rates and cancer network robustness.

Especially, we extended the reachability property to network decomposition and found that *R-core* decomposition is better than *K-core* decomposition in identifying cores that reflect the association between HC entropy and cancer survivability. In addition, it is known that cancer biomarker genes often reside at the innermost core in a biomolecule network. By highlighting examples of other known biomarkers, we showed that the genes with the highest HC values among the innermost *r*-core may be promising candidate cancer biomarkers. Altogether, the proposed approach may be useful in predicting the efficacy of cancer treatment and in identifying putative representative cancer genes.

Supporting information

S1 Fig. The distributions of HC values in pancreatic cancer and basal cell carcinoma signaling networks.
(PDF)

S1 Table. Previously reported cancer biomarker genes among the top three genes with the highest HC values in *1-innermost* R-core.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: Tien-Dzung Tran.

Formal analysis: Tien-Dzung Tran.

Funding acquisition: Tien-Dzung Tran.

Investigation: Tien-Dzung Tran.

Methodology: Tien-Dzung Tran.

Project administration: Yung-Keun Kwon.

Supervision: Yung-Keun Kwon.

Writing – original draft: Tien-Dzung Tran.

Writing – review & editing: Yung-Keun Kwon.

References

1. Thun MJ, DeLancey JO, Center MM, Jemal A, Ward EM. The global burden of cancer: priorities for prevention. Carcinogenesis. 2010; 31(1):100–10. https://doi.org/10.1093/carcin/bgp263 PMID: 19934210

2. Voliotis M, Perrett RM, McWilliams C, McArdle CA, Bowsher CG. Information transfer by leaky, hetero-geneous, protein kinase signaling systems. Proc Natl Acad Sci U S A. 2014; 111(3):E326–33. https://doi.org/10.1073/pnas.1314446111 PMID: 24395805

3. Acencio ML, Bovolenta LA, Camilo E, Lemke N. Prediction of oncogenic interactions and cancer-related signaling networks based on network topology. PLoS One. 2013; 8(10):e77521. https://doi.org/10.1371/journal.pone.0077521 PMID: 24204854

4. Roy J, Winter C, Isik Z, Schroeder M. Network information improves cancer outcome prediction. Brief Bioinform. 2014; 15(4):612–25. https://doi.org/10.1093/bib/bbs083 PMID: 23255167

5. Barjaktarovic Z, Anastasov N, Azimzadeh O, Sriharshan A, Sarioglu H, Ueffing M, et al. Integrative proteomic and microRNA analysis of primary human coronary artery endothelial cells exposed to low-dose gamma radiation. Radiat Environ Biophys. 2013; 52(1):87–98. https://doi.org/10.1007/s00411-012-0439-4 PMID: 23138885

6. Tian T, Olson S, Whitacre JM, Harding A. The origins of cancer robustness and evolvability. Integr Biol (Camb). 2011; 3(1):17–30. https://doi.org/10.1039/C0IB00046A

7. Tran TD, Kwon YK. The relationship between modularity and robustness in signalling networks. J R Soc Interface. 2013; 10(88):20130771. https://doi.org/10.1098/rsif.2013.0771 PMID: 24047877

8. Breitkreutz D, Hlatky L, Rietman E, Tuszynski JA. Molecular signaling network complexity is correlated with cancer patient survivability. Proc Natl Acad Sci U S A. 2012; 109(23):9209–12. https://doi.org/10.1073/pnas.1201416109 PMID: 22615392

9. Takemoto K, Kihara K. Modular organization of cancer signaling networks is associated with patient sur-vivability. Biosystems. 2013; 113(3):149–54. https://doi.org/10.1016/j.biosystems.2013.06.003 PMID: 23810782

10. Janssen BJ, Huizinga EG, Raaijmakers HC, Roos A, Daha MR, Nilsson-Ekdahl K, et al. Structures of complement component C3 provide insights into the function and evolution of immunity. Nature. 2005; 437(7058):505–11. https://doi.org/10.1038/nature04005 PMID: 16177781

11. Leicht EA, Newman ME. Community structure in directed networks. Phys Rev Lett. 2008; 100 (11):118703. https://doi.org/10.1103/PhysRevLett.100.118703 PMID: 18517839

12. Reichardt J, Bornholdt S. Detecting fuzzy community structures in complex networks with a Potts model. Phys Rev Lett. 2004; 93(21):218701. https://doi.org/10.1103/PhysRevLett.93.218701 PMID: 15601068

13. Liu W, Pellegrini M, Wang X. Detecting communities based on network topology. Sci Rep. 2014; 4:5739. https://doi.org/10.1038/srep05739 PMID: 25033828

14. Verma T, Russmann F, Araujo NA, Nagler J, Herrmann HJ. Emergence of core-peripheries in networks. Nat Commun. 2016; 7:10441. https://doi.org/10.1038/ncomms10441 PMID: 26822856

15. Estrada E. Topological structural classes of complex networks. Phys Rev E Stat Nonlin Soft Matter Phys. 2007; 75(1 Pt 2):016103. https://doi.org/10.1103/PhysRevE.75.016103 PMID: 17358220

16. Zhao J, Ding GH, Tao L, Yu H, Yu ZH, Luo JH, et al. Modular co-evolution of metabolic networks. BMC Bioinformatics. 2007; 8:311. https://doi.org/10.1186/1471-2105-8-311 PMID: 17723146

17. Isaac AE, Sinha S. Analysis of core-periphery organization in protein contact networks reveals groups of structurally and functionally critical residues. J Biosci. 2015; 40(4):683–99. https://doi.org/10.1007/s12038-015-9554-0 PMID: 26564971

18. Weng L, Du J, Zhou Q, Cheng B, Li J, Zhang D, et al. Identification of cyclin B1 and Sec62 as biomark-ers for recurrence in patients with HBV-related hepatocellular carcinoma after surgical resection. Mol Cancer. 2012; 11(1):39. https://doi.org/10.1186/1476-4598-11-39 PMID: 22682366

19. Rossa FD, Dercole F, Piccardi C. Profiling core-periphery network structure by random walkers. Sci Rep. 2013; 3:1467. https://doi.org/10.1038/srep01467 PMID: 23507984

20. Ahmed HA, Bhattacharyya DK, Kalita JK. Core and peripheral connectivity based cluster analysis over PPI network. Comput Biol Chem. 2015; 59 Pt B:32–41. https://doi.org/10.1016/j.compbiolchem.2015.08.008 PMID: 26362299

21. Zhang X, Martin T, Newman ME. Identification of core-periphery structure in networks. Phys Rev E Stat Nonlin Soft Matter Phys. 2015; 91(3):032803. https://doi.org/10.1103/PhysRevE.91.032803 PMID: 25871153

22. Batagelj V, Zaversnik M. An O(m) Algorithm for Cores Decomposition of Networks. Computing Research Repository. 2003;cs.DS/0310049.

23. Emerson AI, Andrews S, Ahmed I, Azis TK, Malek JA. K-core decomposition of a protein domain co-occurrence network reveals lower cancer mutation rates for interior cores. Journal of Clinical Bioinformatics. 2015; 5(1):1. https://doi.org/10.1186/s13336-015-0016-6 PMID: 25767694

24. Malliaros FD, Rossi ME, Vazirgiannis M. Locating influential nodes in complex networks. Sci Rep. 2016; 6:19307. https://doi.org/10.1038/srep19307 PMID: 26776455

25. Liu JG, Ren ZM, Guo Q, Chen DB. Evolution characteristics of the network core in the Facebook. PLoS One. 2014; 9(8):e104028. https://doi.org/10.1371/journal.pone.0104028 PMID: 25165852

26. Lu YY, Chen QL, Guan Y, Guo ZZ, Zhang H, Zhang W, et al. Transcriptional profiling and co-expression network analysis identifies potential biomarkers to differentiate chronic hepatitis B and the caused cirrhosis. Mol Biosyst. 2014; 10(5):1117–25. https://doi.org/10.1039/c3mb70474b PMID: 24599568

27. Zhang Y, Li Z, Yang M, Wang D, Yu L, Guo C, et al. Identification of GRB2 and GAB1 coexpression as an unfavorable prognostic factor for hepatocellular carcinoma by a combination of expression profile and network analysis. PLoS One. 2013; 8(12):e85170. https://doi.org/10.1371/journal.pone.0085170 PMID: 24391994

28. Tran TD, Kwon YK. Hierarchical closeness efficiently predicts disease genes in a directed signaling network. Comput Biol Chem. 2014; 53PB:191–197. https://doi.org/10.1016/j.compbiolchem.2014.08.023 PMID: 25462327

29. Fumiã HF, Martins ML. Boolean Network Model for Cancer Pathways: Predicting Carcinogenesis and Targeted Therapy Outcomes. PLOS ONE. 2013; 8(7):1–11.

30. Zhu P, Aliabadi HM, Uludağ H, Han J. Identification of Potential Drug Targets in Cancer Signaling Pathways using Stochastic Logical Models. Scientific Reports. 2016; 6:23078. https://doi.org/10.1038/srep23078 PMID: 26988076

31. Arakelyan A, Nersisyan L. KEGGParser: parsing and editing KEGG pathway maps in Matlab. Bioinformatics. 2013; 29(4):518–9. https://doi.org/10.1093/bioinformatics/bts730 PMID: 23292739

32. Gfeller D, Chappelier JC, De Los Rios P. Finding instabilities in the community structure of complex networks. Phys Rev E Stat Nonlin Soft Matter Phys. 2005; 72(5 Pt 2):056135. https://doi.org/10.1103/PhysRevE.72.056135 PMID: 16383716

33. Waclaw B, Bogacz L, Burda Z, Janke W. Condensation in zero-range processes on inhomogeneous networks. Phys Rev E Stat Nonlin Soft Matter Phys. 2007; 76(4 Pt 2):046114. https://doi.org/10.1103/PhysRevE.76.046114 PMID: 17995067

34. Bianconi G. The entropy of randomized network ensembles. Epl. 2008; 81(2). https://doi.org/10.1209/0295-5075/81/28005

35. Bianconi G, Coolen AC, Perez Vicente CJ. Entropies of complex networks with hierarchically constrained topologies. Phys Rev E Stat Nonlin Soft Matter Phys. 2008; 78(1 Pt 2):016114. https://doi.org/10.1103/PhysRevE.78.016114 PMID: 18764027

36. Bianconi G, Pin P, Marsili M. Assessing the relevance of node features for network structure. Proc Natl Acad Sci U S A. 2009; 106(28):11433–8. https://doi.org/10.1073/pnas.0811511106 PMID: 19571013

37. Banerji CRS, Miranda-Saavedra D, Severini S, Widschwendter M, Enver T, Zhou JX, et al. Cellular network entropy as the energy potential in Waddington's differentiation landscape. Scientific Reports. 2013; 3. https://doi.org/10.1038/srep03039 PMID: 24154593

38. West J, Bianconi G, Severini S, Teschendorff AE. Differential network entropy reveals cancer system hallmarks. Sci Rep. 2012; 2:802. https://doi.org/10.1038/srep00802 PMID: 23150773

39. Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks. 2010; 32(3):245–251. https://doi.org/10.1016/j.socnet.2010.03.006

40. Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, et al. Identification of influential spreaders in complex networks. Nat Phys. 2010; 6(11):888–893. https://doi.org/10.1038/nphys1746

41. Bansal S, Khandelwal S, Meyers LA. Exploring biological network structure with clustered random networks. BMC Bioinformatics. 2009; 10:405. https://doi.org/10.1186/1471-2105-10-405 PMID: 20003212

42. Cranford JA, McCabe SE, Boyd CJ, Lange JE, Reed MB, Scott MS. Effects of residential learning communities on drinking trajectories during the first two years of college. J Stud Alcohol Drugs Suppl. 2009; (16):86–95. https://doi.org/10.15288/jsads.2009.s16.86 PMID: 19538916

43. Bannwarth S, Ait-El-Mkadem S, Chaussenot A, Genin EC, Lacas-Gervais S, Fragaki K, et al. A mitochondrial origin for frontotemporal dementia and amyotrophic lateral sclerosis through CHCHD10 involvement. Brain. 2014; 137(Pt 8):2329–45. https://doi.org/10.1093/brain/awu138 PMID: 24934289

44. Robinson LF, Wager TD, Lindquist MA. Change point estimation in multi-subject fMRI studies. Neuroimage. 2010; 49(2):1581–92. https://doi.org/10.1016/j.neuroimage.2009.08.061 PMID: 19733671

45. Ciliberti S, Martin OC, Wagner A. Robustness can evolve gradually in complex regulatory gene networks with varying topology. PLoS Comput Biol. 2007; 3(2):e15. https://doi.org/10.1371/journal.pcbi.0030015 PMID: 17274682

46. Kitano H. Biological robustness. Nat Rev Genet. 2004; 5(11):826–37. https://doi.org/10.1038/nrg1471 PMID: 15520792

47. Kwon YK, Cho KH. Quantitative analysis of robustness and fragility in biological networks based on feedback dynamics. Bioinformatics. 2008; 24(7):987–94. https://doi.org/10.1093/bioinformatics/btn060 PMID: 18285369

48. Maslov S, Sneppen K. Specificity and Stability in Topology of Protein Networks. Science. 2002; 296 (5569):910–913. https://doi.org/10.1126/science.1065103 PMID: 11988575

49. Maslov S, Sneppen K, Alon U. 8. In: Correlation profiles and motifs in complex networks. Wiley-Blackwell; 2005. p. 168–198. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/3527602755.ch8.

50. Holme P. Metabolic robustness and network modularity: a model study. PLoS One. 2011; 6(2):e16605. https://doi.org/10.1371/journal.pone.0016605 PMID: 21311770

51. Iliopoulos D, Guler G, Han SY, Johnston D, Druck T, McCorkell KA, et al. Fragile genes as biomarkers: epigenetic control of WWOX and FHIT in lung, breast and bladder cancer. Oncogene. 2005; 24 (9):1625–1633. https://doi.org/10.1038/sj.onc.1208398 PMID: 15674328

52. Willem P, Brown J, Schouten J. A novel approach to simultaneously scan genes at fragile sites. BMC Cancer. 2006; 6(1):205. https://doi.org/10.1186/1471-2407-6-205 PMID: 16895604

53. Balasenthil S, Chen N, Lott ST, Chen J, Carter J, Grizzle WE, et al. A migration signature and plasma biomarker panel for pancreatic adenocarcinoma. Cancer Prev Res (Phila). 2011; 4(1):137–49. https://doi.org/10.1158/1940-6207.CAPR-10-0025

54. Jin G, Kim MJ, Jeon HS, Choi JE, Kim DS, Lee EB, et al. PTEN mutations and relationship to EGFR, ERBB2, KRAS, and TP53 mutations in non-small cell lung cancers. Lung Cancer. 2010; 69(3):279–83. https://doi.org/10.1016/j.lungcan.2009.11.012 PMID: 20018398

55. Riely GJ, Marks J, Pao W. KRAS mutations in non-small cell lung cancer. Proc Am Thorac Soc. 2009; 6 (2):201–5. https://doi.org/10.1513/pats.200809-107LC PMID: 19349489

56. Dumenil C, Vieira T, Rouleau E, Antoine M, Duruisseaux M, Poulot V, et al. Is there a specific phenotype associated with the different subtypes of KRAS mutations in patients with advanced non-small-cell lung cancers? Lung Cancer. 2015; 90(3):561–7. https://doi.org/10.1016/j.lungcan.2015.10.012 PMID: 26520186

57. Gaughan EM, Cryer SK, Yeap BY, Jackman DM, Costa DB. Family history of lung cancer in never smokers with non-small-cell lung cancer and its association with tumors harboring EGFR mutations. Lung Cancer. 2013; 79(3):193–7. https://doi.org/10.1016/j.lungcan.2012.12.002 PMID: 23273562

58. Mao CQ, Xiong MH, Liu Y, Shen S, Du XJ, Yang XZ, et al. Synthetic lethal therapy for KRAS mutant non-small-cell lung carcinoma with nanoparticle-mediated CDK4 siRNA delivery. Mol Ther. 2014; 22 (5):964–73. https://doi.org/10.1038/mt.2014.18 PMID: 24496383

59. Rosell R, Moran T, Queralt C, Porta R, Cardenal F, Camps C, et al. Screening for Epidermal Growth Factor Receptor Mutations in Lung Cancer. New England Journal of Medicine. 2009; 361(10):958–967. https://doi.org/10.1056/NEJMoa0904554 PMID: 19692684

60. Garrido-Castro AC, Felip E. HER2 driven non-small cell lung cancer (NSCLC): potential therapeutic approaches. Translational Lung Cancer Research. 2013; 2(2):122–127. https://doi.org/10.3978/j.issn.2218-6751.2013.02.02 PMID: 25806223

61. Longo DM, Louie B, Mathi K, Pos Z, Wang E, Hawtin RE, et al. Racial differences in B cell receptor signaling pathway activation. Journal of Translational Medicine. 2012; 10:113–113. https://doi.org/10.1186/1479-5876-10-113 PMID: 22672557

62. Wiencke JK. Impact of race/ethnicity on molecular pathways in human cancer. Nat Rev Cancer. 2004; 4 (1):79–84. https://doi.org/10.1038/nrc1257 PMID: 14708028

63. Bae NC, Chae MH, Lee MH, Kim KM, Lee EB, Kim CH, et al. EGFR, ERBB2, and KRAS mutations in Korean non-small cell lung cancer patients. Cancer Genetics and Cytogenetics. 2007; 173(2):107–113. https://doi.org/10.1016/j.cancergencyto.2006.10.007 PMID: 17321325

64. Netto GJ. Molecular biomarkers in urothelial carcinoma of the bladder: are we there yet? Nat Rev Urol. 2012; 9(1):41–51.

65. Sathe A, Koshy N, Schmid SC, Thalgott M, Schwarzenbock SM, Krause BJ, et al. CDK4/6 Inhibition Controls Proliferation of Bladder Cancer and Transcription of RB1. J Urol. 2016; 195(3):771–9. https://doi.org/10.1016/j.juro.2015.08.082 PMID: 26318986

66. Zhang T, Wang J, Zhai X, Li H, Li C, Chang J. MiR-124 retards bladder cancer growth by directly targeting CDK4. Acta Biochim Biophys Sin (Shanghai). 2014; 46(12):1072–9. https://doi.org/10.1093/abbs/gmu105

67. Ishii N, Maier D, Merlo A, Tada M, Sawamura Y, Diserens AC, et al. Frequent co-alterations of TP53, p16/CDKN2A, p14ARF, PTEN tumor suppressor genes in human glioma cell lines. Brain Pathol. 1999; 9(3):469–79. https://doi.org/10.1111/j.1750-3639.1999.tb00536.x PMID: 10416987

68. Büschges R, Weber RG, Actor B, Lichter P, Collins VP, Reifenberger G. Amplification and Expression of Cyclin D Genes (CCND1 CCND2 and CCND3) in Human Malignant Gliomas. Brain Pathology. 1999; 9(3):435–442. https://doi.org/10.1111/j.1750-3639.1999.tb00532.x

69. Leslie KK, Sill MW, Lankes HA, Fischer EG, Godwin AK, Gray H, et al. Lapatinib and potential prognostic value of EGFR mutations in a Gynecologic Oncology Group phase II trial of persistent or recurrent endometrial cancer. Gynecologic Oncology. 2012; 127(2):345–350. https://doi.org/10.1016/j.ygyno.2012.07.127 PMID: 22885469

70. Hayes MP, Douglas W, Ellenson LH. Molecular alterations of EGFR and PIK3CA in uterine serous carcinoma. Gynecologic Oncology. 2009; 113(3):370–373. https://doi.org/10.1016/j.ygyno.2008.12.021 PMID: 19272638

71. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, et al. Mutations of the BRAF gene in human cancer. Nature. 2002; 417(6892):949–54. https://doi.org/10.1038/nature00766 PMID: 12068308

72. Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat JP, et al. A landscape of driver mutations in melanoma. Cell. 2012; 150(2):251–63. https://doi.org/10.1016/j.cell.2012.06.024 PMID: 22817889

73. Cicenas J, Tamosaitis L, Kvederaviciute K, Tarvydas R, Staniute G, Kalyan K, et al. KRAS, NRAS and BRAF mutations in colorectal cancer and melanoma. Med Oncol. 2017; 34(2):26. https://doi.org/10.1007/s12032-016-0879-9 PMID: 28074351

74. Krauthammer M, Kong Y, Ha BH, Evans P, Bacchiocchi A, McCusker JP, et al. Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma. Nature Genetics. 2012; 44:1006. https://doi.org/10.1038/ng.2359 PMID: 22842228

75. McCarron SL, Bateman AC, Theaker JM, Howell WM. EGF +61 gene polymorphism and susceptibility to and prognostic markers in cutaneous malignant melanoma. Int J Cancer. 2003; 107(4):673–5. https://doi.org/10.1002/ijc.11448 PMID: 14520709

76. Narang V, Ramli MA, Singhal A, Kumar P, de Libero G, Poidinger M, et al. Automated Identification of Core Regulatory Genes in Human Gene Regulatory Networks. PLoS Comput Biol. 2015; 11(9): e1004504. https://doi.org/10.1371/journal.pcbi.1004504 PMID: 26393364