

RESEARCH

Ranking Cancer Drivers via Betweenness-based Outlier Detection and Random Walks

Cesim Erten¹, Aissa Houdjedj² and Hilal Kazan^{1*}

*Correspondence:

hilal.kazan@antalya.edu.tr

¹Department of Computer Engineering, Antalya Bilim University, Antalya, Turkey

Full list of author information is available at the end of the article

Abstract

Background: Recent cancer genomic studies have generated detailed molecular data on a large number of cancer patients. A key remaining problem in cancer genomics is the identification of driver genes. **Results:** We propose BetweenNet, a computational approach that integrates genomic data with a protein-protein interaction network to identify cancer driver genes. BetweenNet utilizes a measure based on betweenness centrality on patient specific networks to identify the so-called *outlier genes* that correspond to dysregulated genes for each patient. Setting up the relationship between the mutated genes and the outliers through a bipartite graph, it employs a random-walk process on the graph, which provides the final prioritization of the mutated genes. We compare BetweenNet against state-of-the-art cancer gene prioritization methods on lung, breast, and pan-cancer datasets. **Conclusions:** Our evaluations show that BetweenNet is better at recovering known cancer genes based on multiple reference databases. Additionally, we show that the GO terms and the reference pathways enriched in BetweenNet ranked genes and those that are enriched in known cancer genes overlap significantly when compared to the overlaps achieved by the rankings of the alternative methods.

Keywords: Driver gene prioritization, bipartite graph, betweenness centrality, network diffusion

1 Background

Cancer is a complex disease arising in many cases from the effects of multiple genetic changes that give rise to pathway dysregulation through alterations in copy number, DNA methylation, gene expression, and molecular function [1, 2]. Recent cancer genomics projects such as The Cancer Genome Atlas (TCGA) have created a comprehensive catalog of somatic mutations across all major cancer types. A key current challenge in cancer genomics is to distinguish driver mutations that are causal for cancer progression from passenger mutations that do not confer any selective advantage. Consequently, several computational methods have been proposed for the identification of cancer driver genes or driver modules of genes by integrating mutations data with various other types of genetic data [3, 4, 5, 6, 7, 8, 9, 10]; see [11, 12, 13, 14] for recent comprehensive evaluations and surveys on the topic.

Rather than outputting a set of candidate driver genes or modules, a subclass of cancer driver identification methods output a prioritized list of genes ranked by their cancer driving potential. Early approaches in this group have utilized the mutation frequency of each gene by comparing with background mutation rates [15, 16, 17]. However with a careful review of the existing cancer catalogues it

is easy to observe that most tumors share only a small portion of the set of all mutated genes, giving rise to the so called *tumor heterogeneity problem*; methods solely based on mutation rates suffer from low sensitivity due to the existence of long-tail of infrequently mutated genes [4, 18].

One strategy that aims to tackle the long-tail phenomenon is to move from a mutation-centric point of view to a *guilt by association* viewpoint where a correlation between differentially expressed genes and mutated genes are sought. This strategy assumes that even though different sets of genes are mutated in different patients, each of the candidate driver mutations tends to affect a large number of differentially expressed genes. Masica and Karchin present one of the early models based on such a strategy by employing statistical methods for setting up the correlation between mutated genes and the differentially expressed genes to identify candidate drivers [1]. Many different models follow a similar trail by further incorporating biological pathway/network information for setting up such a correlation [6, 19, 20, 21, 22, 23]. DriverNet is among the notable approaches employing mutations data in addition to gene expression and biological network data [19]. It prioritizes mutated genes based on their degrees of network connectivity to dysregulated genes in tumor samples where dysregulation is determined via differential gene expression. Many subsequent approaches are inspired by DriverNet [20, 21, 22, 23]. Among them DawnRank [20], the algorithm by Shi et al. [21], and Subdyquency [23] employ, on top of the overall DriverNet model, versions of heat diffusion on the networks integrating data in the form of biological interactions, mutations, and gene expression. Heat diffusion is a technique employed commonly in many cancer driver gene or gene module discovery algorithms [9, 24, 25, 26, 27, 28]. It generally serves two purposes simultaneously. On the one hand, since the employed interactions data is usually erroneous, diffusing any type of information through the network of interactions, fixes any potential issues arising from missing links in the network. On the other hand, via the diffusion process, it is possible to observe the extent of an effect such as mutation frequency of a gene, at various distant loci in the network. LNDriver extends the DriverNet concept by taking into account gene lengths of the mutated genes to filter out genes that are mutated with high probability due to their lengths [22]. It should be noted that DawnRank and Subdyquency differ slightly from other approaches; the former can identify patient-specific candidate drivers and the latter employs subcellular localization information in addition to the data made of use in the other methods. There are other driver gene prioritization methods that deviate from the overall guilt by association framework, but nevertheless employ different types of genetic data together with the mutations data. IntDriver utilizes an interaction network and gene ontology data within a matrix factorization framework [29]. Dopazo and Erten employ paired data to generate tumor and normal interaction networks filtered with mutations and gene expression data, and measure the efficacy of various graph-theoretical measures in prioritizing breast cancer genes [6]. Note that among the discussed methods DawnRank also utilizes paired data, both from the tumor and the normal samples.

We propose BetweenNet algorithm for cancer driver gene prioritization. Similar to the methods proposed in [20, 21, 22, 23], BetweenNet is also inspired by the DriverNet framework in that it relates the mutated genes and the so-called *outlier*

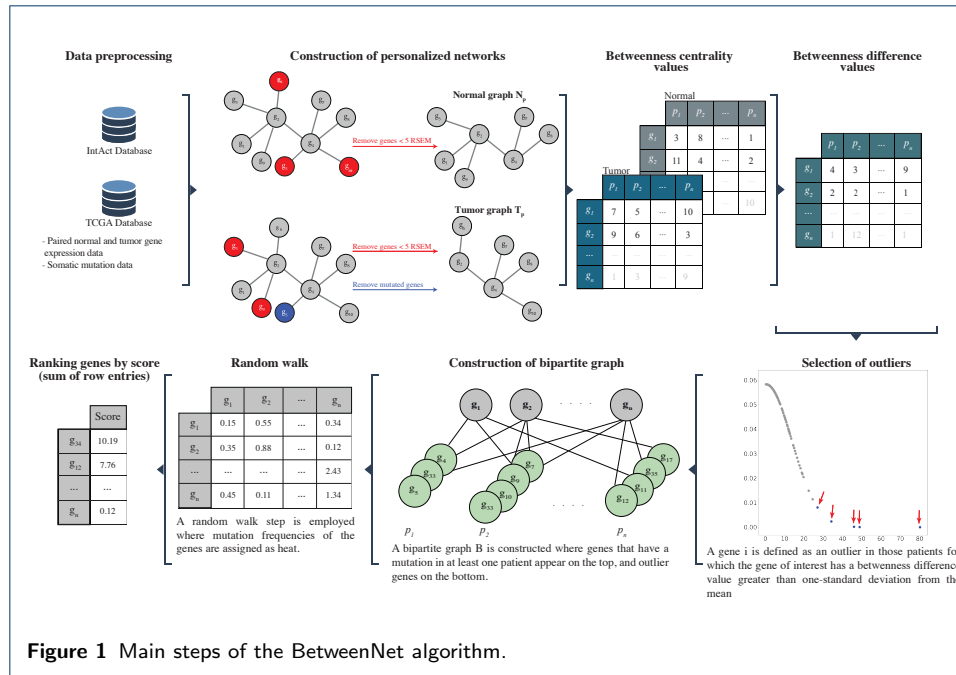


Figure 1 Main steps of the BetweenNet algorithm.

genes corresponding to the dysregulated genes in each patient through a bipartite influence graph. However different from DriverNet and the previous other methods based on it, BetweenNet determines outlier genes based on the betweenness centrality values of the genes in personalized networks. A second contribution of BetweenNet is the employment of a random-walk process on the resulting influence bipartite graph. Random-walks have been utilized in this context previously [21, 23]. However our application of random walk with restart on the whole influence graph is quite different from the two-step or three-step employment of the diffusion process on a per patient basis described in these methods. Through extensive evaluations we demonstrate that BetweenNet outperforms the alternative methods in recovering known reference genes and in providing functionally coherent rankings when compared to the enriched GO terms or the enriched known functional pathways.

2 Methods

We describe the details of the main steps of the BetweenNet algorithm in this section. Figure 1 provides an overview of the algorithm.

2.1 Input data sets and data preparation

In order to construct the pan-cancer cohort, we first identify the cancer types that have more than 10 paired measurements from normal and tumor samples in the TCGA cohort [30] (Supplementary Table 1). We then take the union of all the samples from these cancer types to form the cohort. In addition to the pan-cancer data, we perform separate evaluations on two cancer types. These are breast cancer (BRCA) with 112 samples, lung cancer (LUSC + LUAD) with 109 samples. We download the gene expression (RSEM normalized values [31]) and somatic mutation data for these patients from the Firebrowse database (<http://firebrowse.org>; version 2016_01_28). We exclude the silent mutations in the calculation of mutation

frequencies. In addition to the gene expression and mutations data, we also employ protein-protein interactions data which we gather from the *H. Sapiens* PPI network of the IntAct database [32]. We preprocess the PPI network so that multiple ids for the same protein are merged to a single id. The resulting network contains 16,513 nodes and 189,520 edges.

2.2 Construction of personalized networks

Let $G = (V, E)$ represent the reference *H. Sapiens* PPI network where each vertex $u_i \in V$ denotes a gene i whose expression gives rise to the corresponding protein in the network. Each undirected edge $(u_i, u_j) \in E$ denotes the interaction among the proteins corresponding to the genes i, j . Let P represent the set of patient samples. For each patient $p \in P$, we define two graphs N_p and T_p that represent the PPI networks of the normal and tumor samples, respectively. To construct N_p , we start with the reference PPI network G and remove the nodes that correspond to the genes that are not expressed in the normal sample of the patient p . We deem genes with < 5 RSEM as not expressed. To construct T_p , we remove two sets of genes: (i) genes with < 5 RSEM in the tumor sample; (ii) genes that contain non-silent mutations in the tumor sample.

2.3 Calculation of betweenness centrality values

The standard definition of the betweenness centrality ignores the length of a shortest path. Since considering very long paths as functional relations may not be biologically meaningful, we use a variant of the betweenness centrality called *k-betweenness*, where only shortest paths of length $\leq k$ are included in the calculations [33]. Given an unweighted graph $G = (V, E)$, *k-betweenness* value of a node that corresponds to gene i is defined as follows:

$$\sum_{\forall s, t \in V, s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}} \quad (1)$$

where σ_{st} is the number of shortest paths of length $\leq k$ between genes s and t , and $\sigma_{st}(i)$ is the number of such paths that pass through gene i . We choose k to be 3 in this study and utilize the algorithm presented in Brandes *et al.* to efficiently calculate the *k-betweenness* values [34]. Let $B_{p,i}^N$ and $B_{p,i}^T$ denote the 3-betweenness centrality values of the gene i in the N_p and T_p graphs of the patient p , respectively. We define $B_{p,i}^{diff}$ as $|B_{p,i}^N - B_{p,i}^T|$.

2.4 Selection of outlier genes

For each gene i , we plot the $B_{p,i}^{diff}$ values across all the patients. We observe that the distribution can be approximated with a truncated normal distribution (Supplementary Figure 1). We use the *truncnorm* function in Python to estimate the mean and standard deviation of the distribution. A gene i is defined as an *outlier* in patient p , if $B_{p,i}^{diff}$ is greater than one standard deviation from the mean. We repeat this process for each gene and construct a set of outlier genes for each patient.

2.5 Construction of the bipartite graph

Similar to DriverNet, we construct a bipartite graph B that models the relationship between the set of mutated genes and the outliers. The *mutations partition* of the bipartite graph consists of the genes that have a mutation in at least one patient and the *outliers partition* consists of the outlier genes of all the patients in the cohort. Note that a gene j can be an outlier for multiple patients. In such a case, each occurrence of a gene is represented with a distinct node in the outliers partition of B . Assuming j is an outlier gene for patient p , let u_j^p be the node corresponding to it in the outliers partition. For a node u_i in the mutations partition, edge (u_i, u_j^p) is inserted in B , if gene i is mutated in p and (u_i, u_j) is an edge in G .

2.6 Random walk on the bipartite graph

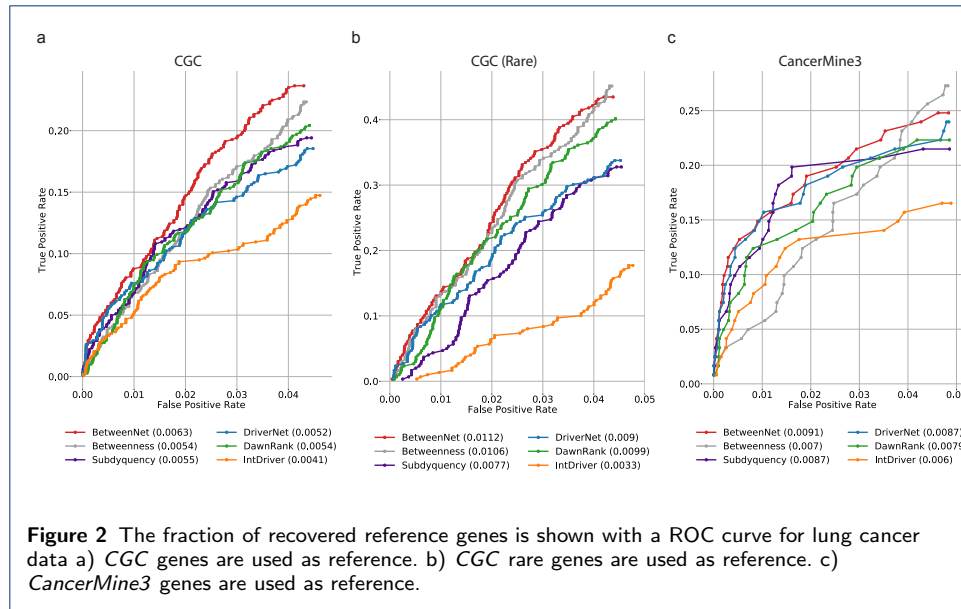
We apply a random walk on the bipartite graph B . The mutation frequencies of the genes are assigned as initial heat values to be diffused throughout the network during random walk. Let $MF(i)$ denote the mutation frequency of gene i , that is, the number of patients where i has a non-silent mutation divided by the total number of patients. Note that heat values are assigned to genes on both sides of the bipartite graph. The random walk starts at a node u_i in B and at each time step moves to one of u_i 's neighbors with probability $1 - \beta$ ($0 \leq \beta \leq 1$). The walk can also restart from u_i with probability β , called the *restart probability*. This process can be defined by a transition matrix T which is constructed by setting $T_{ij} = \frac{1}{deg(u_j)}$ if $(u_i, u_j) \in E$, and $T_{ij} = 0$ otherwise. Here, $deg(u_j)$ corresponds to the degree of the node u_j . Thus T_{ij} can be interpreted as the probability that a simple random walk will transition from u_j to u_i . The random walk process can also be considered as a network propagation process by the equation, $F_{t+1} = (1 - \beta)TF_t + \beta F_0$, where F_t is the distribution of walkers after t steps and F_0 is the diagonal matrix with initial heat values, that is $F_0[i, i] = MF(i)$. We compute the final distribution of the walk by calculating the F matrix iteratively until convergence. We set $\beta = 0.4$ as in previous studies where the effect of varying β has been found to be not significant [9, 28]. After convergence, genes are prioritized by the sum of incoming edge weights.

2.7 Compiling reference gene sets

We compile known cancer genes from the databases Cancer Gene Census (CGC) [35] and CancerMine [36]. From the former, we obtain the list of 728 genes that are found to be associated with cancer. CancerMine uses text-mining to catalogue cancer associated genes where it also extracts information about the type of the cancer. We compile two lists of genes that have at least 3 and 5 citations, respectively. Hereafter, these two reference gene sets are named *CancerMine3* and *CancerMine5*. The number of genes in *CancerMine3* and *CancerMine5* reference sets for each cancer type (i.e., lung, breast) and for pan-cancer cohort are available in the Supplementary Tables 2-4. For lung cancer, we are unable to use *CancerMine5* as a reference due to its small size.

2.8 Enrichment analysis with Gene Ontology and pathway databases

For Gene Ontology (GO) [37] term analysis, we use *goatools*. We download *go-basic.obo* file from <http://geneontology.org/docs/download-ontology/> on June 26th

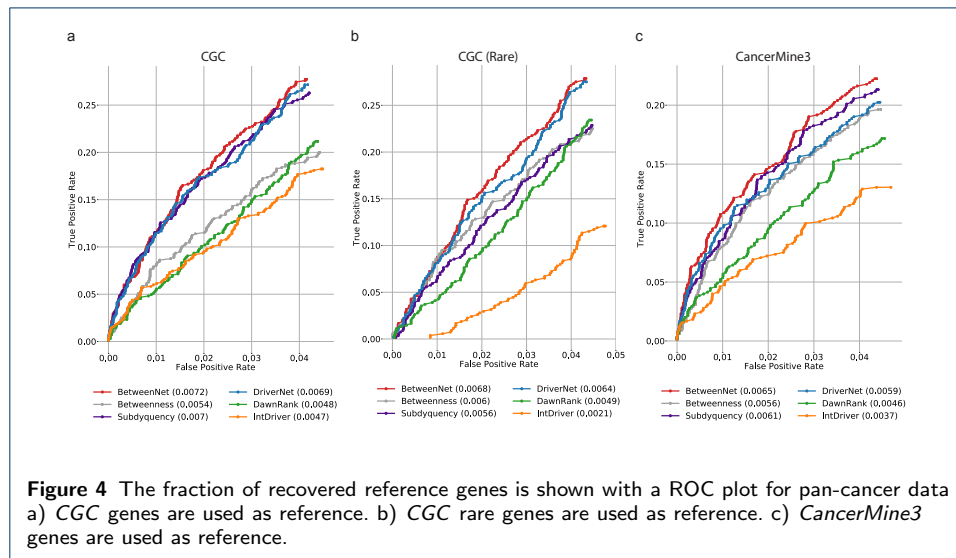
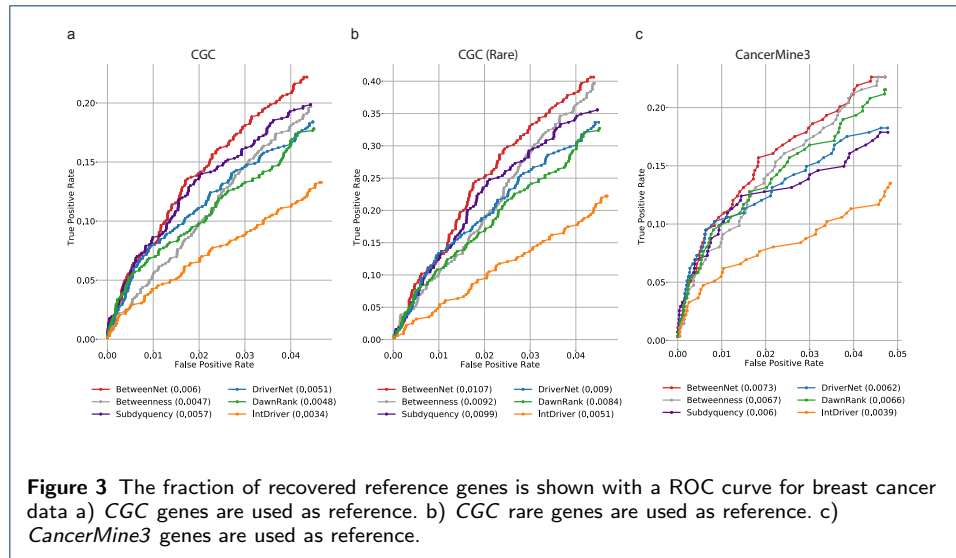


of 2019. We restrict the gene annotations to level 5 by ignoring the higher-level annotations and replacing the deeper-level category annotations with their ancestors at the restricted level.

For the pathway analysis, we use the *AllEnricher* tool with Reactome and Kyoto Encyclopedia of Genes and Genomes (KEGG) [38] pathways. Both *goatools* and *AllEnricher* use Fisher's exact test to calculate p-values and False Discovery Rate (FDR) for multiple testing correction. We use 0.05 as the p-value cutoff to determine significant enrichments.

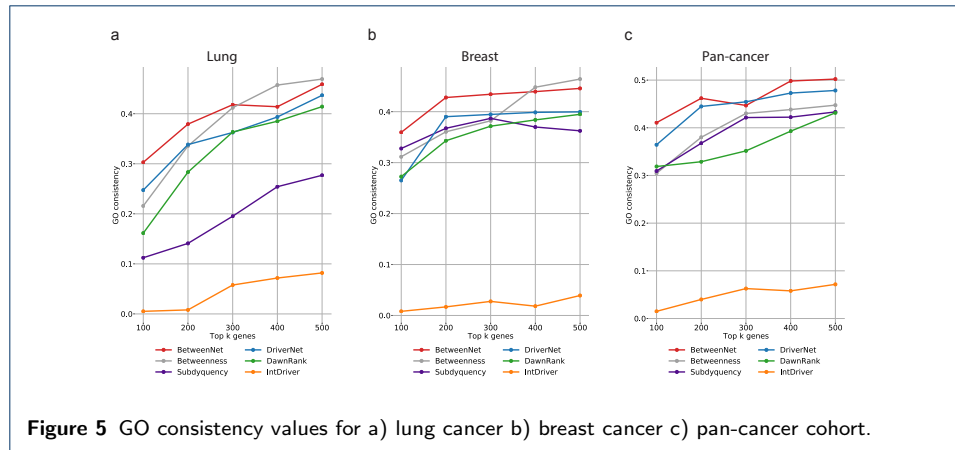
3 Results

We implemented the betweenness centrality measurement algorithm in C++ using the *LEDA* (Library of Efficient Data types and Algorithms) library. The remaining steps are implemented in Python using *NetworkX* library. All the code and necessary datasets are available at <https://github.com/abu-compbio/BetweenNET>. We compare BetweenNet results against those of five other existing cancer driver prioritization methods: DriverNet, Subdyquency, DawnRank, IntDriver, and Dopazo and Erten's prioritization method based on betweenness centrality values, hereafter named only *Betweenness*. Note that for the *Betweenness* method, although the original method ranks all genes, here we only rank mutated genes using the same method for a fair comparison, since all the other methods under consideration are designed to rank mutated genes only. DriverNet is chosen due to its close connection to our work. DawnRank and Subdyquency are included as they extend and improve over DriverNet. *Betweenness* is included as a baseline since our method utilizes a variation of betweenness differences in identifying outlier genes. Finally, IntDriver is included to represent the performance of a distinct strategy that is based on matrix factorization. We evaluate the methods with three datasets: lung cancer, breast cancer, and pan-cancer samples.



3.1 Evaluations with respect to reference cancer gene sets

We first compare the methods based on their ability to recover the sets of known cancer genes. For this, we compute true positive and false positive rates for the top 1000 genes and calculate the area under the ROC (AUROC). Figure 2 shows the ROCs obtained from lung cancer data. In Figure 2-a all CGC genes are used as reference, whereas in Figure 2-b genes with mutation frequencies $\leq 2\%$, namely the *rare drivers*, are included. Figure 2-c is obtained with CancerMine3 as the reference set. BetweenNet achieves a higher AUROC value than all the alternatives for all three reference sets, though the improved performance is more prominent for the first two reference sets. With the CGC reference set, the second ranked method is Subdyquency. This is followed by Betweenness and DawnRank with similar performance with respect to each other. On the other hand, for the CGC-rare reference set, Betweenness and DawnRank perform considerably better than the other three methods. DawnRank's good performance is in line with its focus on finding rare



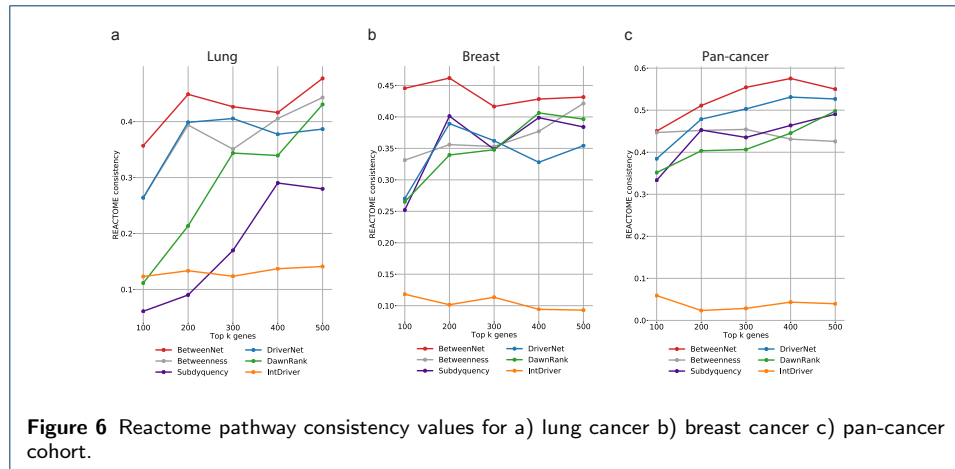
drivers by identifying drivers in a personalized fashion rather than using the mutation frequency of a gene across all the patients. Interestingly, Betweenness results in poor performance with the CancerMine3 reference set where it ranks the second worst method. For all three reference sets, the IntDriver method performs the worst. Overall, these results illustrate the superiority of BetweenNet as it can find both rare and common drivers more accurately.

Figure 3 depicts analogous results for the breast cancer data. As before, BetweenNet achieves the top performance with all three reference sets. For the CGC reference set, Subdyquency and DriverNet rank the second and third, respectively. Betweenness and DawnRank have a similar performance which is worse than Subdyquency and DriverNet. Finally, IntDriver provides the worst performance. For CGC-rare we see a similar performance of the methods except that Betweenness ranks third after BetweenNet and Subdyquency. For CancerMine3, BetweenNet ranks the best which is followed by DawnRank and Betweenness. Results with respect to the CancerMine5 reference set are quite similar and are available in the Supplementary Figure 2-a.

Lastly, Figure 4 shows the results with respect to the pan-cancer dataset. BetweenNet's gene ranking results in the highest AUROC in all three cases, though the difference between BetweenNet and the second best performing method is more prominent for the CancerMine3 reference set. For the CGC reference gene set, BetweenNet, Subdyquency, and DriverNet perform significantly better than the other three methods. For the CGC-rare reference set, DawnRank's performance gets much closer to that of the leading methods, as we previously observe for the other datasets. For the CancerMine3 set, DriverNet and Betweenness can be considered as the second ranking group of methods, whereas DawnRank and IntDriver perform significantly worse than the others. Results with CancerMine5 are similar to those of CancerMine3 and are available in the Supplementary Figure 2-b.

3.2 Evaluations based on functional and pathway analysis

Reference cancer driver gene sets might be incomplete and biased. As such, rather than only finding exact matches between the output gene sets and the reference gene sets, we also define other metrics that measure how well the associated functions of the genes of the two sets match. One such metric is based on GO consistency



(GOC) and the other is based on pathway information. For the former, we find the GO terms enriched in the output gene sets and in the reference gene sets, and check whether the corresponding GO terms overlap. The underlying assumption is that the reference cancer genes and the predicted cancer genes should have similar biological functions. We find the enriched GO terms in the ranked gene sets of varying total sizes from 100 to 500 in the increments of 100 for each method under consideration. We repeat the same GO term enrichment analysis with the reference gene set. We then compute the GOC value between the enriched GO terms of the ranked gene set and those of the reference set, which is defined as the ratio between the size of the intersection of the two sets and the size of the union [39]. Figure 5 shows the GOC values calculated for each cancer type and pan-cancer cohort. We observe that BetweenNet ranked genes perform the best for almost all total size values. For lung cancer, BetweenNet ranks the best in three out of five total size values, whereas Betweenness performs the best in the remaining two cases. DriverNet and Subdyquency rank the third and fourth, respectively. This is followed by DawnRank and IntDriver with notably large differences between the performances of these two methods. For breast cancer, BetweenNet performs significantly better than the other methods. All the rest of the methods except IntDriver exhibit similar performances whereas IntDriver performs significantly worse than the others. For pan-cancer data, the methods can be ranked as follows: BetweenNet, DriverNet, Betweenness, Subdyquency, DawnRank, and IntDriver.

We repeat the same type of analysis with pathways as well, this time replacing GO term enrichment with pathway enrichment. Namely, we identify the pathways enriched in the reference set of genes and the set of genes output by a ranking method. We then compute the number of pathways common in both of these sets. Figure 6 shows the results with Reactome reference pathways for all cancer types. For all cases, BetweenNet outperforms the other methods with a large margin. For lung cancer, the second best method is Betweenness, whereas for breast and pan-cancer it is difficult to identify the second best method as it varies for each total size value. For breast and pan-cancer, IntDriver performs worse than the other methods. Results obtained with KEGG look similar and can be found in the supplementary.

Analysis of BetweenNet ranked genes

We further explore BetweenNet's top 20 ranking genes for each cancer type and identify those that do not appear in the cancer reference gene set CGC. We investigate the literature for existing work that study these genes within the context of cancer biology. For lung cancer, the gene ranked the sixth by BetweenNet, the Vascular cell adhesion molecule-1 (VCAM-1) gene has been found to have higher expression in cancer-associated fibroblasts and blocking it attenuates the proliferation and invasion of lung cancer cells [40]. VCAM-1 can easily be missed by the models solely based on mutation frequency, as it is mutated in only three patients across the lung cancer cohort. Indeed, only BetweenNet ranks this gene within the top 20. Another gene that ranks in our top 20 for lung cancer is Ryanodine receptor 2 (RYR2). This gene is found to be strongly associated with a subtype of lung adenocarcinoma that shows high tumor mutational burden [41]. RYR2 mutations were also found to be significantly higher in non-small cell lung cancer patients that live in highly polluted regions as compared to control regions [42]. Lastly, TP53BP1 is another gene that is only identified by BetweenNet within the top 20. TP53BP1 is a critical component of DNA damage response (DDR) machinery and defects of this machinery are involved in multiple processes of tumorigenesis [43, 44, 45]. A recent study showed that TP53BP1 protein expression is significantly reduced in several cancer tissues including lung carcinoma suggesting a role as tumor suppressor [46]. FN1 is within the top 20 genes of the Betweenness ranking for breast cancer. A recent study confirms that increased fibronectin is linked to metastasis through its role in facilitation of cancer cell attachment and spreading in 3D suspension culture of MDA-MB-231 cell line [47]. As such, existing studies indicate that FN1 is not a driver but play critical roles in metastasis which can still be of interest for therapeutic purposes. Another top ranking gene for breast cancer is GABARAP. GABARAP has been shown to involve in tumorigenesis *in vivo* by delaying cell death and its associated immune-related response [48].

In pan-cancer data, CDC5L, VCAM1, TP53BP1, FN1, MKI67, and PLEC are among the top 20 ranking genes that do not appear in CGC. For each of these genes, there are multiple studies in the literature indicating roles for cancer development and progression. Additionally, many of these genes also appear in top 20 for individual cancer types we study, that is in lung and breast cancer. For instance, CDC5L is a transcription factor that is associated with tumorigenesis in multiple cancer types including prostate, hepatocellular, colorectal, and bladder cancer [49, 50, 51, 52]. VCAM1 is linked to tumor immune evasion in more than ten cancer types [53, 54, 55, 56, 57, 58, 59, 60], TP53BP1 also appears in our *CancerMine3* list and its role in cancer development is recently reviewed in [61]. There are studies showing FN1's aberrant expression in multiple cancer types [62, 63, 64, 65]. Lastly, MKI67 expression has been shown as a biomarker for multiple cancer types [66, 67, 68, 69, 70].

Discussion

Having shown that BetweenNet performs better than existing methods, we finally investigate whether all components of BetweenNet contribute to this performance by testing against simpler alternatives. We test a version of BetweenNet where the

random walk step is omitted. In this version, there are no edge weights and ranking is simply based on the degrees of the bipartite influence graph nodes corresponding to the mutated genes. We also include DriverNet in this comparison to evaluate the effects of the outlier selection process. Comparing the version of BetweenNet without the random walk step against DriverNet reveals the added value of defining outlier genes based on betweenness centrality values rather than gene expression. Supplementary Figures 4-6 compare these three models on lung, breast, and pan-cancer data. Except for experiments with *CancerMine3* in lung cancer and *CGC* in pan-cancer, we observe that BetweenNet achieves the top performance for all cancer types and for all reference gene sets.

Conclusions

We propose BetweenNet, a novel cancer driver gene prioritization approach that integrates genomic data with the connectivity within PPI networks. One contribution of BetweenNet is the identification of patient specific dysregulated genes with a measure based on betweenness centrality on personalized networks. BetweenNet ranks mutated genes by their effects on dysregulated genes. To characterize these effects, a bipartite influence graph is formed to represent the relations between the mutated genes and dysregulated genes in each patient. Another contribution of BetweenNet is the employment of a random-walk process on the resulting influence bipartite graph. Through careful comparisons, we show that both the use of betweenness centrality metric and the employment of random walk have added values in identification of cancer driver genes. We also demonstrate that BetweenNet outperforms the alternative methods in recovering known reference genes and in providing functionally coherent rankings with three large-scale TCGA datasets: lung cancer, breast cancer, and pan-cancer samples. Additionally, we find that many of our top ranking genes that do not appear in reference cancer gene sets have roles in cancer development based on existing literature. Taken together, our results indicate that BetweenNet effectively integrates genomic data and connectivity information to prioritize cancer driver genes.

Abbreviations

GO: Gene Ontology, CGC: Cancer Gene Census, TCGA: The Cancer Genome Atlas, PPI: Protein-Protein Interaction, BRCA: Breast Invasive Carcinoma LUAD: Lung Adenocarcinoma, LUSC: Lung Squamous Cell Carcinoma, LEDA: Library of Efficient Data Types and Algorithms, AUROC: Area Under the Receiver Operating Characteristics, GOC: GO consistency.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The source code and datasets used in this research can be downloaded from <https://github.com/abu-compbio/BetweenNET>

Competing interests

The authors declare that they have no competing interests.

Funding

This work has been supported by the Scientific and Technological Research Council of Turkey [117E879 to H.K. and C.E.].

Authors' contributions

Authors are listed in alphabetical order with respect to last names. CE and HK designed the overall framework. CE, AH, and HK contributed equally in the algorithm design. CE and HK designed the evaluation framework. AH implemented the algorithm, the evaluation scripts, and the performance tests. CE, AH, and HK contributed equally in writing the manuscript.

Acknowledgements

Not applicable.

Author details

¹Department of Computer Engineering, Antalya Bilim University, Antalya, Turkey. ² Electrical and Computer Engineering Graduate Program, Antalya Bilim University, Antalya, Turkey.

References

1. Masica, D., Karchin, R.: Correlation of somatic mutation and expression identifies genes important in human glioblastoma progression and survival. *Cancer research* **71**, 4550–61 (2011)
2. Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Kinzler, K.W.: Cancer genome landscapes. *Science (New York, N.Y.)* **339**(6127), 1546–1558 (2013)
3. Erten, S., Bebek, G., Koyuturk, M.: Vavien: An algorithm for prioritizing candidate disease genes based on topological similarity of proteins in interaction networks. *J Comput Biol* **18**, 1561–74 (2011). doi:10.1089/cmb.2011.0154
4. Lawrence, M., Stojanov, P., Polak, P., Kryukov, G., Cibulskis, K., Sivachenko, A., Carter, S., Stewart, C., H Mermel, C., Roberts, S., Kiezun, A., Hammerman, P., McKenna, A., Drier, Y., Zou, L., Ramos, A., Pugh, T., Stransky, N., Helman, E., Getz, G.: Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499** (2013). doi:10.1038/nature12213
5. Yang, H., Wei, Q., Zhong, X., Yang, H., Li, B.: Cancer driver gene discovery through an integrative genomics approach in a non-parametric bayesian framework. *Bioinformatics* **33**(4), 483–490 (2017)
6. Dopazo, J., Erten, C.: Graph-theoretical comparison of normal and tumor networks in identifying BRCA genes. *BMC Systems Biology* **11**(1) (2017). doi:10.1186/s12918-017-0495-0
7. Miller, C.A., Settle, S.H., Sulman, E.P., Aldape, K.D., Milosavljevic, A.: Discovering functional modules by identifying recurrent and mutually exclusive mutational patterns in tumors. *BMC Medical Genomics* **4**(1), 34 (2011). doi:10.1186/1755-8794-4-34
8. Vandin, F., Upfal, E., Raphael, B.J.: *De Novo* discovery of mutated driver pathways in cancer. In: Research in Computational Molecular Biology - 15th Annual International Conference, RECOMB 2011, Vancouver, BC, Canada, March 28-31, 2011. Proceedings, pp. 499–500 (2011). doi:10.1007/978-3-642-20036-6_44. https://doi.org/10.1007/978-3-642-20036-6_44
9. Leiserson, M.D.M., Vandin, F., Wu, H.-T., Dobson, J.R., Eldridge, J.V., Thomas, J.L., Papoutsaki, A., Kim, Y., Niu, B., McLellan, M., Lawrence, M.S., Gonzalez-Perez, A., Tamborero, D., Cheng, Y., Ryslik, G.A., Lopez-Bigas, N., Getz, G., Ding, L., Raphael, B.J.: Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nature Genetics* **47**(2), 106–114 (2014). doi:10.1038/ng.3168
10. Liu, B., Wu, C., Shen, X., Pan, W.: A novel and efficient algorithm for de novo discovery of mutated driver pathways in cancer. *Ann. Appl. Stat.* **11**(3), 1481–1512 (2017). doi:10.1214/17-AOAS1042
11. Dimitrakopoulos, C.M., Beerenwinkel, N.: Computational approaches for the identification of cancer genes and pathways. Wiley. Published online 11 November 2016. (2017). doi:10.1002/wsbm.1364
12. Zhang, J., Zhang, S.: The discovery of mutated driver pathways in cancer: Models and algorithms. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **15**(3), 988–998 (2018). doi:10.1109/TCBB.2016.2640963
13. Bailey, M., Tokheim, C., Porta, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M., Kim, J., Reardon, B., Ng, K.s., Jeong, K., Cao, S., Wang, Z., Gao, J., Gao, Q., Wang, F., Liu, E.M., Mularoni, L., Mariamidze, A.: Comprehensive characterization of cancer driver genes and mutations. *Cell* **173**, 371–38518 (2018). doi:10.1016/j.cell.2018.02.060
14. Tokheim, C., Papadopoulos, N., Kinzler, K., Vogelstein, B., Karchin, R.: Evaluating the evaluation of cancer driver genes. *Proceedings of the National Academy of Sciences* **113**, 201616440 (2016). doi:10.1073/pnas.1616440113
15. Dees, N.D., Zhang, Q., Kandoth, C., Wendl, M.C., Schierding, W., Koboldt, D.C., Mooney, T.B., Callaway, M.B., Dooling, D., Mardis, E.R., Wilson, R.K., Ding, L.: MuSiC: Identifying mutational significance in cancer genomes. *Genome Research* **22**(8), 1589–1598 (2012). doi:10.1101/gr.134635.111
16. Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., Kiezun, A., Hammerman, P.S., McKenna, A., Drier, Y., Zou, L., Ramos, A.H., Pugh, T.J., Stransky, N., Helman, E., Kim, J., Sougnez, C., Ambrogio, L., Nickerson, E., Shefler, E., Cortés, M.L., Auclair, D., Saksena, G., Voet, D., Noble, M., Dicara, D., Lin, P., Lichtenstein, L., Heiman, D.I., Fennell, T., Imielinski, M., Hernandez, B., Hodis, E., Baca, S., Dulak, A.M., Lohr, J., Landau, D.A., Wu, C.J., Melendez-Zajgla, J., Hidalgo-Miranda, A., Koren, A., McCarroll, S.A., Mora, J., Lee, R.S., Crompton, B., Onofrio, R., Parkin, M., Winckler, W., Ardlie, K., Gabriel, S.B., Roberts, C.W.M., Biegel, J.A., Stegmaier, K., Bass, A.J., Garraway, L.A., Meyerson, M., Golub, T.R., Gordenin, D.A., Sunyaev, S., Lander, E.S., Getz, G.: Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**(7457), 214–218 (2013). doi:10.1038/nature12213
17. Hodis, E., Watson, I.R., Kryukov, G.V., Arol, S.T., Imielinski, M., Theurillat, J.P., Nickerson, E., Auclair, D., Li, L., Place, C., Dicara, D., Ramos, A.H., Lawrence, M.S., Cibulskis, K., Sivachenko, A., Voet, D., Saksena, G., Stransky, N., Onofrio, R.C., Winckler, W., Ardlie, K., Wagle, N., Wargo, J., Chong, K., Morton, D.L., Stemke-Hale, K., Chen, G., Noble, M., Meyerson, M., Ladbury, J.E., Davies, M.A., Gershenwald, J.E., Wagner,

- S.N., Hoon, D.S.B., Schadendorf, D., Lander, E.S., Gabriel, S.B., Getz, G., Garraway, L.A., Chin, L.: A landscape of driver mutations in melanoma. *Cell* **150**(2), 251–263 (2012). doi:10.1016/j.cell.2012.06.024
18. Leiserson, M.D.M., Blokh, D., Sharan, R., Raphael, B.J.: Simultaneous identification of multiple driver pathways in cancer. *PLOS Comput Biol* **9**(5), 1–15 (2013). doi:10.1371/journal.pcbi.1003054
 19. Bashashati, A., Haffari, G., Ding, J., Ha, G., Lui, K., Rosner, J., Huntsman, D.G., Caldas, C., Aparicio, S.A., Shah, S.P.: DriverNet: uncovering the impact of somatic driver mutations on transcriptional networks in cancer. *Genome Biology* **13**(12), 124 (2012). doi:10.1186/gb-2012-13-12-r124
 20. Hou, J.P., Ma, J.: Dawnrank: discovering personalized driver genes in cancer. *Genome Medicine* **6**(56), 1–16 (2014)
 21. Shi, K., Gao, L., Wang, B.: Discovering potential cancer driver genes by an integrated network-based approach. *Mol. BioSyst.* **12** (2016). doi:10.1039/C6MB00274A
 22. Wei, P.-J., Zhang, D., Xia, J., Zheng, C.-H.: Lndriver: Identifying driver genes by integrating mutation and expression data based on gene-gene interaction network. *BMC Bioinformatics* **17**, 221–230 (2016). doi:10.1186/s12859-016-1332-y
 23. Song, J., Peng, W., Wang, F.: A random walk-based method to identify driver genes by integrating the subcellular localization and variation frequency into bipartite graph. *BMC Bioinformatics* **20**(1), 238 (2019). doi:10.1186/s12859-019-2847-9
 24. Vandin, F., Upfal, E., Raphael, B.: Algorithms for detecting significantly mutated pathways in cancer. *J Comput Biol* **18**, 507–22 (2011). doi:10.1089/cmb.2010.0265
 25. Bersanelli, M., Mosca, E., Remondini, D., Castellani, G., Milanese, L.: Network diffusion-based analysis of high-throughput data for the detection of differentially enriched modules. *Scientific Reports* **6**(1), 34841 (2016). doi:10.1038/srep34841
 26. Yang, C., Ge, S.-G., Zheng, C.-H.: ndmaSNF: cancer subtype discovery based on integrative framework assisted by network diffusion model. *Oncotarget* **8**(51), 89021–89032 (2017). doi:10.18632/oncotarget.21643
 27. Reyna, M.A., Leiserson, M.D.M., Raphael, B.J.: Hierarchical HotNet: identifying hierarchies of altered subnetworks. *Bioinformatics* **34**(17), 972–980 (2018)
 28. Ahmed, R., Baali, I., Erten, C., Hoxha, E., Kazan, H.: MEXCOwalk: Mutual Exclusion and Coverage Based Random Walk to Identify Cancer Modules. *Bioinformatics* **36**(3), 872–879 (2020)
 29. Xi, J., Wang, M., Li, A.: Discovering potential driver genes through an integrated model of somatic mutation profiles and gene functional information. *Mol. BioSyst.* **13** (2017). doi:10.1039/C7MB00303J
 30. Tomczak, K., Czerwińska, P., Wiznerowicz, M.: The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemporary oncology (Poznan, Poland)* **19**(1A), 68–77 (2015). doi:10.5114/wo.2014.47136
 31. Li, B., Dewey, C.N.: RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**(1), 323 (2011). doi:10.1186/1471-2105-12-323
 32. Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., Campbell, N.H., Chavali, G., Chen, C., del-Toro, N., Duesbury, M., Dumousseau, M., Galeota, E., Hinz, U., Iannuccelli, M., Jagannathan, S., Jimenez, R., Khadake, J., Lagreid, A., Licata, L., Lovering, R.C., Meldal, B., Melidoni, A.N., Milagros, M., Peluso, D., Perfetto, L., Porras, P., Raghunath, A., Ricard-Blum, S., Roechert, B., Stutz, A., Tognolli, M., van Roey, K., Cesareni, G., Hermjakob, H.: The mintact project—intact as a common curation platform for 11 molecular interaction databases. *Nucleic acids research* **42**(Database issue), 358–63 (2014). doi:10.1093/nar/gkt1115
 33. Borgatti, S.P., Everett, M.G.: A graph-theoretic perspective on centrality. *Social Networks* **28**(4), 466–484 (2006)
 34. Brandes, U.: On variants of shortest-path betweenness centrality and their generic computation. *Social Networks* **30**(2), 136–145 (2008). doi:10.1016/j.socnet.2007.11.001
 35. Forbes, S.A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., Cole, C.G., Ward, S., Dawson, E., Ponting, L., Stefancsik, R., Harsha, B., Kok, C.Y., Jia, M., Jubb, H., Sondka, Z., Thompson, S., De, T., Campbell, P.J.: COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Research* **45**(D1), 777–783 (2017). doi:10.1093/nar/gkw1121
 36. Lever, J., Zhao, E.Y., Grewal, J., Jones, M.R., Jones, S.J.M.: CancerMine: a literature-mined resource for drivers, oncogenes and tumor suppressors in cancer. *Nature methods* **16**(6), 505–507 (2019). doi:10.1038/s41592-019-0422-y
 37. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., Sherlock, G.: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* **25**(1), 25–9 (2000). doi:10.1038/75556
 38. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., Morishima, K.: KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research* **45**(D1), 353–361 (2017). doi:10.1093/nar/gkw1092
 39. Aladağ, A.E., Erten, C.: SPINAL: scalable protein interaction network alignment. *Bioinformatics* **29**(7), 917–924 (2013). doi:10.1093/bioinformatics/btt071
 40. Zhou, Z., Zhou, Q., Wu, X., Xu, S., Hu, X., Tao, X., Li, B., Peng, J., Li, D., Shen, L., Cao, Y., Yang, L.: VCAM-1 secreted from cancer-associated fibroblasts enhances the growth and invasion of lung cancer cells through AKT and MAPK signaling. *Cancer Letters* **473**, 62–73 (2020). doi:10.1016/j.canlet.2019.12.039
 41. Wang, C., Liang, H., Lin, C., Li, F., Xie, G., Qiao, S., Shi, X., Deng, J., Zhao, X., Wu, K., Zhang, X.: Molecular subtyping and prognostic assessment based on tumor mutation burden in patients with lung adenocarcinomas. *International Journal of Molecular Sciences* **20**(17) (2019). doi:10.3390/ijms20174251
 42. Yu, X.J., Yang, M.J., Zhou, B., Wang, G.Z., Huang, Y.C., Wu, L.C., Cheng, X., Wen, Z.S., Huang, J.Y., Zhang, Y.D., Gao, X.H., Li, G.F., He, S.W., Gu, Z.H., Ma, L., Pan, C.M., Wang, P., Chen, H.B., Hong, Z.P., Wang, X.L., Mao, W.J., Jin, X.L., Kang, H., Chen, S.T., Zhu, Y.Q., Gu, W.Y., Liu, Z., Dong, H., Tian, L.W., Chen, S.J., Cao, Y., Wang, S.Y., Zhou, G.B.: Characterization of Somatic Mutations in Air Pollution-Related Lung Cancer. *EBioMedicine* **2**(6), 583–590 (2015). doi:10.1016/j.ebiom.2015.04.003
 43. Kastan, M.B., Bartek, J.: Cell-cycle checkpoints and cancer (2004). doi:10.1038/nature03097

44. Nevanlinna, H., Bartek, J.: The CHEK2 gene and inherited breast cancer susceptibility (2006). doi:10.1038/sj.onc.1209877
45. Bartek, J., Lukas, J.: DNA damage checkpoints: from initiation to recovery or adaptation (2007). doi:10.1016/j.ceb.2007.02.009
46. Bartkova, J., Hořejší, Z., Sehested, M., Nesland, J.M., Rajpert-De Meyts, E., Skakkebaek, N.E., Stucki, M., Jackson, S., Lukas, J., Bartek, J.: DNA damage response mediators MDC1 and 53BP1: Constitutive activation and aberrant loss in breast and lung cancer, but not in testicular germ cell tumours. *Oncogene* **26**(53), 7414–7422 (2007). doi:10.1038/sj.onc.1210553
47. Park, H.J., Helfman, D.M.: Up-regulated fibronectin in 3D culture facilitates spreading of triple negative breast cancer cells on 2D through integrin β -5 and Src. *Scientific Reports* **9**(1) (2019). doi:10.1038/s41598-019-56276-3
48. Salah, F.S., Ebbinghaus, M., Muley, V.Y., Zhou, Z., Al-Saadi, K.R.D., Pacyna-Gengelbach, M., O'sullivan, G.A., Betz, H., König, R., Wang, Z.Q., Bräuer, R., Petersen, I.: Tumor suppression in mice lacking GABARAP, an atg8/lc3 family member implicated in autophagy, is associated with alterations in cytokine secretion and cell death. *Cell Death and Disease* **7**(4) (2016). doi:10.1038/cddis.2016.93
49. Li, X., Wang, X., Song, W., Xu, H., Huang, R., Wang, Y., Zhao, W., Xiao, Z., Yang, X.: Oncogenic properties of NEAT1 in prostate cancer cells depend on the CDC5L-AGRN transcriptional regulation circuit. *Cancer Research* **78**(15), 4138–4149 (2018). doi:10.1158/0008-5472.CAN-18-0688
50. Qiu, H., Zhang, X., Ni, W., Shi, W., Fan, H., Xu, J., Chen, Y., Ni, R., Tao, T.: Expression and Clinical Role of Cdc5L as a Novel Cell Cycle Protein in Hepatocellular Carcinoma. *Digestive Diseases and Sciences* **61**(3), 795–805 (2016). doi:10.1007/s10620-015-3937-9
51. Li, J., Zhang, N., Zhang, R., Sun, L., Yu, W., Guo, W., Gao, Y., Li, M., Liu, W., Liang, P., Deng, W., Cui, X.: CDC5L Promotes hTERT Expression and Colorectal Tumor Growth. *Cellular Physiology and Biochemistry* **41**(6), 2475–2488 (2017). doi:10.1159/000475916
52. Zhang, Z., Mao, W., Wang, L., Liu, M., Zhang, W., Wu, Y., Zhang, J., Mao, S., Geng, J., Yao, X.: Depletion of CDC5L inhibits bladder cancer tumorigenesis. *Journal of Cancer* **11**(2), 353–363 (2020). doi:10.7150/jca.32850
53. Chen, S., Su, X., Jiang, X., Zhang, T., Min, I., Ding, Y., Wang, X., Mao, Z., Cao, J., Teng, X., Fahey, T.J., Wang, W., Teng, L.: VCAM-1 Upregulation Contributes to Insensitivity of Vemurafenib in BRAF-Mutant Thyroid Cancer. *Translational Oncology* **13**(2), 441–451 (2020). doi:10.1016/j.tranon.2019.10.007
54. Siyasi, M., Mahjoubi, F., Mahjoubi, B., Shabani, S.: Study of VCAM-1 Gene Expression in Normal and Tumoral Tissues in Patients with Colorectal Cancer. *Journal of Biotechnology and Biomedical Science* **1**(1), 19–26 (2017). doi:10.14302/issn.2576-6694.jbbs-17-1744
55. Chang, A.C., Chen, P.C., Lin, Y.F., Su, C.M., Liu, J.F., Lin, T.H., Chuang, S.M., Tang, C.H.: Osteoblast-secreted WISP-1 promotes adherence of prostate cancer cells to bone via the VCAM-1/integrin α 4 β 1 system. *Cancer Letters* **426**, 47–56 (2018). doi:10.1016/j.canlet.2018.03.050
56. Derosa, G., Maffioli, P.: vol. 59(5)
57. Lu, X., Mu, E., Wei, Y., Riethdorf, S., Yang, Q., Yuan, M., Yan, J., Hua, Y., Tiede, B., Lu, X., Haffty, B., Pantel, K., Massagué, J., Kang, Y.: VCAM-1 Promotes Osteolytic Expansion of Indolent Bone Micrometastasis of Breast Cancer by Engaging α 4 β 1-Positive Osteoclast Progenitors. *Cancer Cell* **20**(6), 701–714 (2011). doi:10.1016/j.ccr.2011.11.002
58. Chen, Q., Zhang, X.H.F., Massagué, J.: Macrophage Binding to Receptor VCAM-1 Transmits Survival Signals in Breast Cancer Cells that Invade the Lungs. *Cancer Cell* **20**(4), 538–549 (2011). doi:10.1016/j.ccr.2011.08.025
59. Shioi, K.I., Komiya, A., Hattori, K., Huang, Y., Sano, F., Murakami, T., Nakaigawa, N., Kishida, T., Kubota, Y., Nagashima, Y., Yao, M.: Vascular cell adhesion molecule 1 predicts cancer-free survival in clear cell renal carcinoma patients. *Clinical Cancer Research* **12**(24), 7339–7346 (2006). doi:10.1158/1078-0432.CCR-06-1737
60. Heidemann, J., Maaser, C., Lügering, A., Spahn, T.W., Zimmer, K.P., Herbst, H., Rafiee, P., Domschke, W., Kriegstein, C.F., Binion, D.G., Kucharzik, T.F.: Expression of vascular cell adhesion molecule-1 (CD 106) in normal and neoplastic human esophageal squamous epithelium. *International Journal of Oncology* **28**(1), 77–85 (2006). doi:10.3892/ijo.28.1.77
61. Mirza-Aghazadeh-Attari, M., Mohammadzadeh, A., Yousefi, B., Mihanfar, A., Karimian, A., Majidinia, M.: 53BP1: A key player of DNA damage response with critical functions in cancer. *Elsevier B.V.* (2019). doi:10.1016/j.dnarep.2018.11.008
62. Jerhammar, F., Ceder, R., Garvin, S., Grénman, R., Grafström, R.C., Roberg, K.: Fibronectin 1 is a potential biomarker for radioresistance in head and neck squamous cell carcinoma. *Cancer Biology and Therapy* **10**(12), 1244–1251 (2010). doi:10.4161/cbt.10.12.13432
63. Sponziello, M., Rosignolo, F., Celano, M., Maggisano, V., Pecce, V., De Rose, R.F., Lombardo, G.E., Durante, C., Filetti, S., Damante, G., Russo, D., Bulotta, S.: Fibronectin-1 expression is increased in aggressive thyroid cancer and favors the migration and invasion of cancer cells. *Molecular and Cellular Endocrinology* **431**, 123–132 (2016). doi:10.1016/j.mce.2016.05.007
64. Das, D.K., Naidoo, M., Ilboudo, A., Park, J.Y., Ali, T., Krampis, K., Robinson, B.D., Osborne, J.R., Ogunwobi, O.O.: miR-1207-3p regulates the androgen receptor in prostate cancer via FNDC1/fibronectin. *Experimental Cell Research* **348**(2), 190–200 (2016). doi:10.1016/j.yexcr.2016.09.021
65. Cai, X., Liu, C., Zhang, T.N., Zhu, Y.W., Dong, X., Xue, P.: Down-regulation of FN1 inhibits colorectal carcinogenesis by suppressing proliferation, migration, and invasion. *Journal of Cellular Biochemistry* **119**(6), 4717–4728 (2018). doi:10.1002/jcb.26651
66. Pan, D.H., Wen, D.Y., Luo, Y.H., Chen, G., Yang, H., Chen, J.Q., He, Y.: The diagnostic and prognostic values of Ki-67/ MIB-1 expression in thyroid cancer: A meta-analysis with 6,051 cases. *Dove Medical Press Ltd.* (2017). doi:10.2147/OTT.S135593
67. Jin, Y., Zhang, W., Liu, B.: [Abnormal expression of p53, Ki67 and iNOS in human esophageal carcinoma in situ and pre-malignant lesions]. *Zhonghua zhong liu za zhi [Chinese journal of oncology]* **23**(2), 129–31 (2001)
68. Liu, B., Zhang, J., Huang, C., Liu, H.: Dyskerin Overexpression in Human Hepatocellular Carcinoma Is

- Associated with Advanced Clinical Stage and Poor Patient Prognosis. *PLoS ONE* **7**(8), 43147 (2012). doi:10.1371/journal.pone.0043147
69. Yang, C., Su, H., Liao, X., Han, C., Yu, T., Zhu, G., Wang, X., Winkler, C.A., O'Brien, S.J., Peng, T.: Marker of proliferation Ki-67 expression is associated with transforming growth factor beta 1 and can predict the prognosis of patients with hepatic B virus-related hepatocellular carcinoma. *Cancer Management and Research* **10**, 679–696 (2018). doi:10.2147/CMAR.S162595
70. Ko, K., Jeong, C.W., Kwak, C., Kim, H.H., Ku, J.H.: Significance of Ki-67 in non-muscle invasive bladder cancer patients: A systematic review and meta-analysis. *Oncotarget* **8**(59), 100614–100630 (2017). doi:10.18632/oncotarget.21899