BIOINFORMATICS AND NETWORK MEDICINE

Putative disease gene identification and drug repurposing for NON-SMALL CELL LUNGS CARCINOMA (NSCLC)

Lucia Dicunta, Elisa Pierini, Valeria Popolla

GROUP 3

ABSTRACT

Non-small-cell lung cancer (NSCLC) represents a heterogeneous group of malignancies that are the leading cause of cancer-related death worldwide. Although many NSCLC-related genes and pathways have been identified, there remains an urgent need to mechanistically understand how these genes and pathways drive NSCLC. This work provides an extensive analysis of the networks associated to this disease, starting from the characterization of the subnetwork of the known disease genes within the interactome. Then, three different Disease-Gene Prediction algorithms have been implemented to extract, from the Protein-Protein Interaction (PPI) network, a list of possible candidates to study in order to better understand the mechanisms underlying the disease. The results have been compared, validated computationally by k-fold cross validation and evaluated by performing enrichment analysis. Finally, a Drug-Gene Interaction database has been used to find the most promising genes among the candidates for drug repurposing. This further evaluation shows the differences among the approaches that the different algorithms have and the strength of computational methods in biology. Genes predicted by modularity-based algorithms focus more attention on genes that in biology are more likely to be involved in the disease, the direct physical interacting genes, while diffusion algorithms, even with a small diffusion parameter t, go further and better identify those genes that are not suspected to be involved in the module. However, it seems that the 3 putative disease modules are valid and interpretable. The general aim of the work is giving a support to scientific research and biology to fasten the procedures leading to a better understanding of the disease and therefore, to a more effective and specialized treatment.

INTRODUCTION

Cancer is a heterogeneous disease that refers to a collection of vastly different cellular states with dysregulated cell signaling and regulatory circuits. Among them, lung cancers are the leading cause of cancer-related deaths worldwide. The prognosis for lung cancer is grim, as more than half of individuals diagnosed succumb within a year, and the 5-year survival rate is below 18%. Non-small cell lung cancer (NSCLC) manifestations constitute the majority of cases, with identified risk factors including cigarette smoking, environmental elements, and genetic factors. Advances in genetics and biomarker testing have allowed the identification of specific mutations, enhancing personalized treatment approaches. In the existing literature, different methods have been implemented to identify the disease module related to NSCLC. They differ, in particular, for the definition of the underlying interaction network, the data sources and gathering. For instance, a tentative of characterizing and predicting the disease module (Wang et al.) is based on the construction of the interactome by prioritizing those genes coming out from a RW algorithm. Starting from it, communities have been detected to identify biomarkers and possible drug targets. Another computational analysis has been provided by Wang et al, also focused on community detection among the metastases of NSCLC: the results lead to the definition of t issue-specific (TS) or non-tissue-specific (NTS) modules. While the research of topological modules within the subnetworks associated to the disease is giving insightful results, looking for the overlapping of disease module and topological module within the PPI, with the available knowledge, does not seem promising. In this sense our study confirms the **hub hypothesis**: disease genes segregate at the functional periphery of the interactome. As a matter of fact, Louvain algorithm applied to the network, shows a non-significative distribution of seed genes across the communities detected.

This study employs a comprehensive methodology, utilizing various software resources and libraries. In the 'Materials and Methods' section disease gene identification and drug repurposing methods in non-small cell lung cancer (NSCLC) are investigated. The approach involves reconstructing the interactome, comparative analysis of algorithms, and subsequent enrichment analysis to uncover potential pathways involved in NSCLC. The study also explores drug repurposing by identifying top-ranked drugs associated with putative disease genes. The 'Results and Discussion' section provides insights into the known disease module, algorithm validation outcomes and enriched biological processes related to NSCLC as well as the top-ranked drugs associated with putative disease genes.

MATERIALS AND METHODS

1. Software resources and libraries

Data manipulation, prediction algorithms and the validation procedures have been implemented in Python language. Useful were libraries such as *panda*s and *numpy*, for data gathering and filtering, *networkx*, for network analysis and *DIAMOnD* library for the modularity-based DG prediction. Moreover, Cytoscape and EnrichR have been used, the former for predicting genes by diffusion, the latter for enrichment analysis.

2. PPI and GDA data gathering and interactome reconstruction.

The first step consisted in reconstructing and customizing the interactome, downloaded from from BioGRID latest release. More importance was given to physical association between proteins since these kinds of interactions play a crucial role in various cellular processes and biological functions. In detail, the file containing protein-protein interactions related to all organisms has been downloaded and filtered keeping those regarding 'Homo sapiens' only. We removed all non-human interactions so that 'organism A' and 'organism B' were both equal to 9606 (Homo sapiens) and kept only the type 'physical' interactions. Also self-loops and redundancies have been filtered out (using BioGRID ID as unique indicator). To isolate the largest connected component of the interactome (LCC), we built the related graph from the DataFrame edgelist using as nodes *'Official Symbol Interactor A'* and *'Official Symbol Interactor B'* and then checked the number of connected components. Since there was only one, no further manipulations were necessary.

Gene-disease associations (GDAs) related to NSCLC were gathered from DisGeNET and filtered for 'CURATED'. The resulting DataFrame is made of the genes to which the disease is associated. After that, by overlapping the interactome and the GDA database a list of seed genes has been identified. In this way it was possible to characterize the disease network, the partial and known disease module. An interesting fact regarding this disease can be anticipated: unlike the majority of known disease module (Ghiassian et al.) the NSCLC has in its Largest Connected Component (LCC) more than the 80% of the GDAs. A summary of the GDA-related data is available in the below section 'Results and Discussion' as well as the table about the network measures (Node degree, Betweenness centrality, Eigenvector centrality, Closeness centrality, ratio Betweenness/Node degree) related to the first 50 disease genes in the disease LCC ordered for node degree from higher to lower and the scatterplot representing the relationship between node degree and node betweenness.

3. Comparative analysis of the disease genes identification algorithms

In order to infer and validate putative disease genes, three algorithms were used:

- **DIAMOnD** (DIseAse MOdule Detection). It is based on a systematic analysis of the network properties of known disease genes which reveals that the connectivity significance is the most predictive quantity characterizing their interaction patterns. The algorithm firstly determines the connectivity significance for all genes connected to any of the S_0 seed genes, then ranks genes according to their respective p-values, adds to the set of seed nodes the gene with

the highest rank and repeats these steps with the expanded set of seed genes pulling one gene at a time into the growing disease module until it spans across the entire network.

- **DiaBLE** (Diamond Background Local Expansion). It works similarly to DIAMOnD except for the universe size which is the *smallest universe expansion* of the seed at iteration i. This universe size N_i is a set composed of seed genes, genes that have at least one link to the seed set (candidate genes) and their first neighbors.
- **Cytoscape Diffusion.** It exploits the formalism of *network propagation*, a powerful concept in bioinformatics and, in general, in network science. By running the diffusion algorithm over the PPI network, starting from the known seed genes, a robust estimate of network distance between sets of nodes has been defined to choose the candidate putative genes. In this way, the network of interactions is useful to find new genes that are most relevant to a set of well understood genes.

These three algorithms have been computationally validated using a 5-fold cross validation which consists of splitting the disease genes set S_0 into 5 subsets, selecting, each time, one subset as probe set S_p and the remaining four subsets as training set S_T , running the algorithm using the S_T sets and checking the output for genes in the S_p set. The metrics used to compare the results are **Precision**, fraction of corrected instances among the retrieved instances, **Recall**, true positive rate, and **F1-score**, the summary among the first two metrics.

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

These metrics were computed considering the top 50 positions and the top X positions where $X = \frac{1}{10}n,\ \frac{1}{4}n,\ \frac{1}{2}n,\ n$ and $n =$ number of known GDAs. The results are reported in the 'Results and Discussion' section below.

4. Putative disease gene identification

The best performing algorithm has been chosen to extract a valid list of 100 candidate putative disease genes. With this list of genes, **Enrichment Analysis** has been performed to determine whether certain pathways or annotations are statistically enriched in the putative disease genes and for the original disease genes gathered in the prediction phase compared to what would be expected by chance. We also evaluated the overlap between enriched functions of original disease genes and putative disease genes. The datasets used for enrichment are KEGG and the Gene Ontology, the former to identify the common biological pathways and metabolic processes enriched in the new set of putative genes, the latter for the comparison of the original and predicted genes in the 3 different domains of the ontology (Biological Process, Molecular Function, Cellular Component)

5. Drug repurposing

Once the list of candidate putative genes has been obtained, the top 20 have been chosen to perform a first drug repurposing analysis.

We downloaded from DGIdb, the gene-drug interactions database, and filtered it for APPROVED drugs and for the selected putative disease genes. A ranking of the identified drugs has been generated starting from those associated with the most of the above 20 genes going down. For the first three drugs we checked for clinical trials testing the drug for the NSCLC on https://clinicaltrials.gov and noted the number of clinical trials.

RESULTS AND DISCUSSION

1. Known Disease Module characterization

The customized interactome consists of 19816 genes and 809943 interactions. In this network the disease module containing the know seed genes is represented as a subnetwork of 144 nodes out of the total known set of genes associated to the disease (156 genes). The LCC of the module is made up of 120 genes, almost the whole subnetwork, as shown in **Fig.1.** The disease LCC is represented by the red nodes. Even if the whole network seems to be well connected, it is disconnected overall.

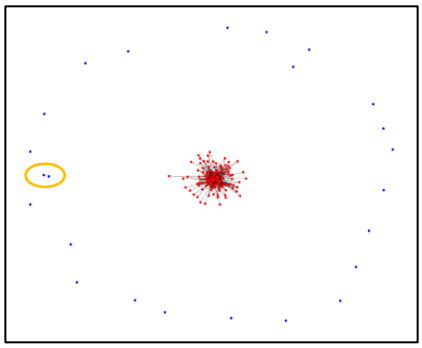

***Fig. 1**: Disease LCC*

As a matter of fact, the 24 genes that are not in the LCC are distributed in 23 connected components: all communities of one node except for one having two connected nodes (circled in orange). This naïve characterization of the disease network does not particularly highlight the modularity properties of the disease within the interactome. Related studies on the disease have shown that, by refining the construction of the interactome, 8 communities can be identified.

The tables below represent the summary of GDAs and the average measures of centrality in the LCC of the disease.

Disease Name	UMLS disease ID	MeSH disease class	Number of associated genes	Number of genes present in the interactome	LCC size of disease interactome
NSCLC	C0007131	C08	156	144	120

***Table 1**. Basic network data*

Degree Centrality	Betweenness Centrality	Eigenvector Centrality	Closeness Centrality	Betweeness/Degree
0.070028	0.012873	0.066483	0.406903	0.183831

***Table 2**. Average Centrality measures in the LCC*

The disease LCC exhibits a network structure where nodes have low degrees of connectivity, with a relatively high average closeness centrality. This means that the diameter of the LCC is quite low, and, despite the low average betweenness centrality, every two nodes in the network are connected by a very short path. However, the presence of nodes with higher betweenness centrality compared to their degree suggests potential key players acting as bridges in the network.

Gene	Degree Centrality	Betweenness Centrality	Eigenvector Centrality	Closeness Centrality	Betweeness/Degree
MYC	51	0.187884	0.333005	0.597990	0.003684
TP53	38	0.103346	0.271803	0.550926	0.002720
…	…	…	…	…	…

***Table 3.** Preview of the centrality measures of the top 50 nodes (more in the appendix)*

The **Table 3.** above shows those metrics for the first 50 disease genes, sorted according to the degree centrality, in the disease LCC. The first ranked gene *(MYT),* with a very high degree centrality, is a common oncogene, a transcription factor with a wide array of functions affecting cellular activities. This gene has a high betweenness centrality with respect to the other nodes in the edge: we assume that it represents a hub in the disease subnetwork. For what concerns the other nodes, notice that lower values of betweenness centrality correspond to lower values f betweenness centrality.

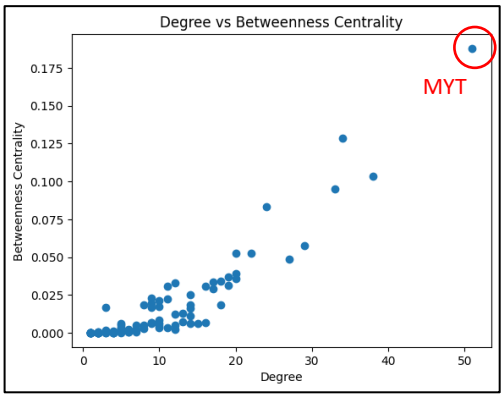

***Fig. 2**. Degree vs Betweenness*

2. Computational Validation

The 5- fold cross validation gives an estimate of the performance of the three algorithms and return an indubitable winner, Cytoscape diffusion algorithm. **Fig.3** represents the trend of the computed metrics as the portion of predicted genes increases. A possible explanation of this result can be given if considering both the nature of the different algorithms and the chosen validation procedure. Keep in mind also the characterization of the disease subnetwork in the previous section.

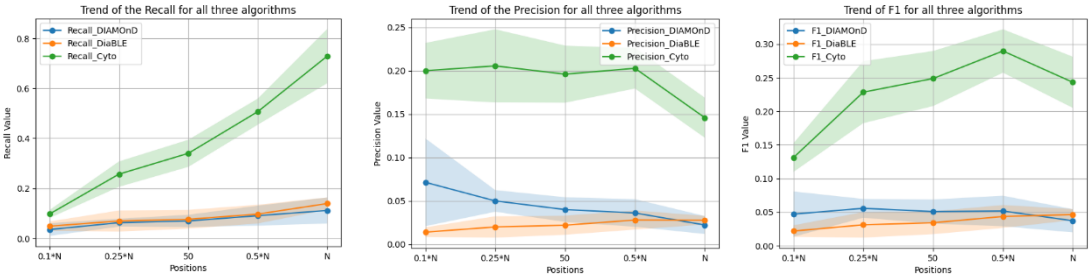

***Fig.3**. Precision, recall and f1-score*

The validation procedure uses in turn a subset of the seed genes. Hence, the metric is strongly affected by the limited knowledge coming from BioGRID. Modularity-based algorithms such as DIAMOnD and DiaBLE, based on the connectivity significance, will identify direct neighbors of the genes in the LCC of the disease as putative nodes, while a distance measure based on a random walk algorithm will *diffuse* beyond the LCC and can also reach the disconnected genes in the module.

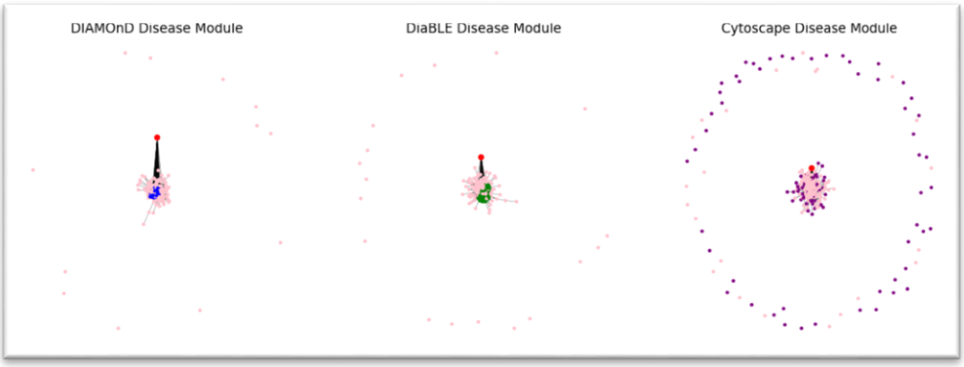

***Fig.4**. DIAMOnD, DiaBLE and Cytoscape putative disease module. Pink nodes are the original seed genes. In red is drawn the only common gene for the 3 algorithms. The distance of the node from the LCC is proportional its degree centrality measure*.

3. Enrichment Analysis

At this point, the diffusion-based algorithm was used to predict 100 new putative disease genes using all known GDAs as seed genes. This list was, then, used for the enrichment analysis as well as the list of the original disease genes gathered.

For what concerns the Biological Process, shared processes have been identified: "**Regulation Of Cell Population Proliferation," "Regulation Of Apoptotic Process,**" and **"Transmembrane Receptor Protein Tyrosine Kinase Signaling Pathway."** These processes play a crucial role in disease contexts, indicating potential involvement in disease pathogenesis. The 'Regulation of Cell Population Proliferation' signifies critical control over cell growth, a fundamental process in tumor formation. Similarly, the 'Regulation of Apoptotic Process' underscores the importance of regulating programmed cell death, while the 'Transmembrane Receptor Protein Tyrosine Kinase Signaling Pathway' highlights the regulation of key signaling pathways involved in cell proliferation and survival.

Then the analysis of Molecular Function has identified crucial activities such as **'Protein Tyrosine Kinase Activity,' 'Growth Factor Receptor Binding,**' and **'Protein Phosphatase Binding**.' 'Protein Tyrosine Kinase Activity' highlights the protein's ability to phosphorylate tyrosine residues, playing a key role in regulating intracellular signaling pathways and modulating cell growth. Similarly, 'Growth Factor Receptor Binding' indicates the protein's affinity for growth factor receptors, implying involvement in the regulation of cell growth and survival. The activity of 'Protein Phosphatase Binding' suggests interaction with enzymes that dephosphorylate proteins, contributing to the delicate balance of phosphorylation and dephosphorylation in cellular signaling pathways.

The Cellular Component describes where the gene or protein functions within the cell, such as the nucleus, cell membrane, or cytoplasm. We have identified two common components across both datasets: "**Focal Adhesion"** and "**Intracellular Membrane-Bounded Organelle."** The protein FAK (Focal Adhesion Kinase) directly regulates many fundamental signaling pathways related to adhesion and growth factors, crucial in the context of human cancer. This suggests that FAK may play a significant role in tumor formation and dissemination. "Intracellular Membrane-Bounded Organelle" indicates the subcellular location of a structure within a cell, precisely organelles surrounded by a membrane. Each organelle performs specific functions within the cell. The annotation provides information about the subcellular position of molecular activities or functions, contributing to understanding the spatial context of cellular activities. The presence or alteration of specific organelles within cells could impact gene regulation, cellular communication, and key processes related to lung carcinogenesis.

About the metabolic pathways from KEGG that are common in seed and candidate genes "**Pathways in cancer**" and the **"PI3K-Akt signaling pathway"** are pertinent for understanding cancer progression and cellular signaling pathways. The subtype "non-small lung carcinoma" is exclusively present in the original dataset, suggesting significant molecular diversity or a specific pathogenesis associated with lung cancer.

4. Drug Repurposing

Even if in validation step results were definitely better for the diffusion algorithm, we found DiaBLE putative genes list more promising with respect to Drug Repurposing task and related studies. As a matter of fact, for Cytoscape candidates, there are few drugs related to only one gene out of the top three, while for DiaBle, 5 candidate genes result related to the first ranked drug, 3 to the second and 3 to the third. It is interesting to notice that not only the majority of the studies are open but also related to the disease of interest. A justification for the consistent number of research for modularity-based putative genes can be given by the fact that the direct neighbors are intuitively more likely to be involved in the disease and therefore more studied. However, the diffusion algorithm, the best performing according to computational validation, returns a valid candidate also according to the associated clinical trials. 20 studies are conducted on VANDETANIB for the treatment of patients suffering from this kind of lung cancer. Not surprisingly the candidate gene has an oncogenic role in many types of tumors. As a conclusion, it can be stated that, despite the differences in the approaches, all the algorithms predict a credible disease module.

AUTHOR CONTRIBUTIONS

In general, all the group members have been actively involved in the conduction of the study by continuous discussions and updates. We decided to not split the work in silos to not lose information on the related topics and to have a broader understanding of the material. Also, the drafting of the report and the notebook is the result of a combination of our ideas, attempts of implementation and personal styles.

REFERENCES

Ghiassian SD, Menche J, Barabási A-L A DIseASe MOdule Detection (DIAMOnD) Algorithm Derived from a Systematic Analysis of Connectivity Patterns of Disease Proteins in the Human Interactome. PLoS Comput Biol , 2015, 11(4): e1004120. doi:10.1371/journal.pcbi.1004120 (https://github.com/dinaghiassian/DIAMOnD)

Petti M, Bizzarri D, Verrienti A, Falcone R, Farina L, Connectivity significance for disease gene prioritization in an expanding universe, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2019, DOI 10.1109/TCBB.2019.2938512

Ahmadi, S.E., Rahimi, S., Zarandi, B. et al. MYC: a multipurpose oncogene with prognostic and therapeutic implications in blood malignancies. J Hematol Oncol 14, 121 (2021). https://doi.org/10.1186/s13045-021-01111-4

Wang F, Han S, Yang J, Hu G, Knowledge-Guided "Community Network" Analysis Reveals the Functional Modules and Candidate Targets in Non-Small-Cell Lung Cancer, Cells, 2021, doi: 10.3390/cells10020402 PMCID: PMC7919838 PMID: 33669233

Wang K, Zhang M, Wang J, Sun P, Luo J, Jin H, Li R, Pan C, Lu L. A Systematic Analysis Identifies Key Regulators Involved in Cell Proliferation and Potential Drugs for the Treatment of Human Lung Adenocarcinoma. Front Oncol. 2021 Sep 28;11:737152. doi: 10.3389/fonc.2021.737152. PMID: 34650921; PMCID: PMC8505978.

Yamaoka, T.; Kusumoto, S.; Ando, K.; Ohba, M.; Ohmori, T. Receptor Tyrosine Kinase-Targeted Cancer Therapy. *Int. J. Mol. Sci.* 2018, *19*, 3491. https://doi.org/10.3390/ijms19113491

Stefan C. Grant, Sally A. Lynch, John Mendelsohn, Growth factors in lung cancer: Current status and implications for therapy, JAI, 1997, https://doi.org/10.1016/S1874-5687(97)80013-4.

Mirzapoiazova T, Xiao G, Mambetsariev B, Nasser MW, Miaou E, Singhal SS, Srivastava S, Mambetsariev I, Nelson MS, Nam A, Behal A, Arvanitis LD, Atri P, Muschen M, Tissot FLH, Miser J, Kovach JS, Sattler M, Batra SK, Kulkarni P, Salgia R. Protein Phosphatase 2A as a Therapeutic Target in Small Cell Lung Cancer. Mol Cancer Ther. 2021 Oct;20(10):1820-1835. doi: 10.1158/1535-7163.MCT-21-0013. Epub 2021 Jul 12. Erratum in: Mol Cancer Ther. 2022 Apr 1;21(4):700. PMID: 34253596; PMCID: PMC8722383.

Zhang, H., Shao, H., Golubovskaya, V. et al. Efficacy of focal adhesion kinase inhibition in non-small cell lung cancer with oncogenically activated MAPK pathways. Br J Cancer 115, 203–211 (2016). https://doi.org/10.1038/bjc.2016.190

Díaz P, Sandoval-Bórquez A, Bravo-Sagua R, Quest A, Lavandero S, Perspectives on Organelle Interaction, Protein Dysregulation, and Cancer Disease, Frontiers in Cell and Developmental Biology , 2021, 10.3389/fcell.2021.613336

Hemmings BA, Restuccia DF. PI3K-PKB/Akt pathway. Cold Spring Harb Perspect Biol. 2012 Sep 1;4(9):a011189. doi: 10.1101/cshperspect.a011189. Erratum in: Cold Spring Harb Perspect Biol. 2015 Apr;7(4). pii: a026609. doi: 10.1101/cshperspect.a026609. PMID: 22952397; PMCID: PMC3428770.

Ahn ER, Mangat PK, Garrett-Mayer E, Halabi S, Dib EG, Haggstrom DE, Alguire KB, Calfa CJ, Cannon TL, Crilley PA, Gaba AG, Marr AS, Sangal A, Thota R, Antonelli KR, Islam S, Rygiel AL, Bruinooge SS, Schilsky RL. Palbociclib in Patients With Non-Small-Cell Lung Cancer With CDKN2A Alterations: Results From the Targeted Agent and Profiling Utilization Registry Study. JCO Precis Oncol. 2020 Nov;4:757-766. doi: 10.1200/PO.20.00037. PMID: 35050752.

Wang G, Bie F, Li G, Shi J, Zeng Y, Du J. Study of the co-expression gene modules of non-small cell lung cancer metastases. Cancer Biomark. 2021;30(3):321-329. doi: 10.3233/CBM-201605. PMID: 33337349.

Zappa C, Mousa SA. Non-small cell lung cancer: current treatment and future advances. Transl Lung Cancer Res. 2016 Jun;5(3):288-300. doi: 10.21037/tlcr.2016.06.07. PMID: 27413711; PMCID: PMC4931124

Ahn ER, Mangat PK, Garrett-Mayer E, Halabi S, Dib EG, Haggstrom DE, Alguire KB, Calfa CJ, Cannon TL, Crilley PA, Gaba AG, Marr AS, Sangal A, Thota R, Antonelli KR, Islam S, Rygiel AL, Bruinooge SS, Schilsky RL. Palbociclib in Patients With Non-Small-Cell Lung Cancer With CDKN2A Alterations: Results From the Targeted Agent and Profiling Utilization Registry Study. JCO Precis Oncol. 2020 Nov;4:757-766. doi: 10.1200/PO.20.00037. PMID: 35050752.

Armando Santoro, Wu-Chou Su, Alejandro Navarro, Matteo Simonelli, James CH Yang, Andrea Ardizzoni, Fabrice Barlesi, Jin Hyoung Kang, Sarah DiDominick, Ahmed Abdelhady, Xueying Chen, Uz Stammberger, Enriqueta Felip, Phase Ib/II study of ceritinib in combination with ribociclib in patients with ALK-rearranged non–small cell lung cancer, Lung Cancer, Volume 166, 2022, Pages 170-177, ISSN 0169-5002.

Naz S, Sowers A, Choudhuri R, et al. Abemaciclib, a Selective CDK4/6 Inhibitor, Enhances the Radiosensitivity of Non-Small Cell Lung Cancer, Clinical Cancer Research : an Official Journal of the American Association for Cancer Research. 2018 Aug;24(16):3994-4005. DOI: 10.1158/1078-0432.ccr-17-3575. PMID: 29716919; PMCID: PMC6137329.

Ma W, Zhu M, Wang B, Gong Z, Du X, Yang T, Shi X, Dai B, Zhan Y, Zhang D, Ji Y, Wang Y, Li S, Zhang Y. Vandetanib drives growth arrest and promotes sensitivity to imatinib in chronic myeloid leukemia by targeting ephrin type-B receptor 4. Mol Oncol. 2022 Jul;16(14):2747-2765. doi: 10.1002/1878-0261.13270. Epub 2022 Jun 27. PMID: 35689424; PMCID: PMC9297786.

APPENDIX

	gene	degree	betweenness_centrality	eigenvector_centrality	closeness_centrality	betweenness_degree_ratio
0	MYC	51	0.187884	0.333005	0.597990	0.003684
1	TP53	38	0.103346	0.271803	0.550926	0.002720
2	EGFR	34	0.128787	0.181036	0.528889	0.003788
3	RAF1	33	0.095293	0.242075	0.540909	0.002888
4	AKT1	29	0.057696	0.235590	0.533632	0.001990
5	BAP1	27	0.048442	0.203356	0.510730	0.001794
6	STAT3	24	0.083169	0.149455	0.515152	0.003465
7	MDM2	22	0.052306	0.168784	0.477912	0.002378
8	ERBB2	20	0.052483	0.148017	0.497908	0.002624
9	RB1	20	0.038938	0.153424	0.493776	0.001947
10	KDM1A	20	0.035551	0.153447	0.485714	0.001778
11	PTEN	19	0.031261	0.154850	0.491736	0.001645
12	HIF1A	19	0.036608	0.155562	0.485714	0.001927
13	SMARCA4	18	0.018154	0.162677	0.483740	0.001009
14	ENO1	18	0.033945	0.142004	0.487705	0.001886
15	SOX2	17	0.033361	0.102640	0.455939	0.001962
16	KRAS	17	0.029091	0.107465	0.459459	0.001711
17	STK11	16	0.006432	0.164125	0.476000	0.000402
18	ITGB1	16	0.030582	0.089893	0.454198	0.001911
19	CDKN2A	15	0.005914	0.160779	0.468504	0.000394
20	MAP2K1	14	0.016123	0.101539	0.450758	0.001152
21	UCHL1	14	0.024877	0.136475	0.476000	0.001777
22	E2F1	14	0.011206	0.119422	0.445693	0.000800
23	CYLD	14	0.018582	0.095390	0.463035	0.001327
24	TERT	14	0.006359	0.130939	0.450758	0.000454
25	NRAS	13	0.012620	0.105388	0.474104	0.000971
26	MET	13	0.007275	0.128354	0.477912	0.000560
27	PPIA	12	0.033092	0.091418	0.445693	0.002758
28	RASSF1	12	0.002313	0.130992	0.454198	0.000193
29	TP53BP1	12	0.012023	0.091853	0.442379	0.001002
30	NFE2L2	12	0.005238	0.111608	0.457692	0.000436
31	CORO1C	11	0.030567	0.087217	0.466667	0.002779
32	MTOR	11	0.022277	0.090739	0.455939	0.002025

33	BRAF	11	0.003114	0.089107	0.421986	0.000283
34	CAT	10	0.021330	0.073439	0.434307	0.002133
35	PRDX1	10	0.003299	0.096926	0.455939	0.000330
36	MAP2K2	10	0.004253	0.072576	0.428058	0.000425
37	SHMT2	10	0.008516	0.082324	0.449057	0.000852
38	CASP8	10	0.017353	0.084558	0.442379	0.001735
39	E2F6	10	0.005950	0.069503	0.413194	0.000595
40	PIK3CA	9	0.022997	0.076999	0.447368	0.002555
41	LGALS1	9	0.006696	0.070607	0.447368	0.000744
42	PHGDH	9	0.006908	0.080790	0.449057	0.000768
43	KEAP1	9	0.019762	0.079134	0.454198	0.002196
44	FGFR1	9	0.016985	0.054837	0.413194	0.001887
45	PXN	9	0.005984	0.061222	0.423488	0.000665
46	LRRC59	8	0.018427	0.070540	0.442379	0.002303
47	ALK	8	0.004924	0.039393	0.397993	0.000615
48	GSTP1	8	0.002677	0.078274	0.440741	0.000335
49	PGAM1	7	0.002409	0.053918	0.408935	0.000344

Table 3. *Preview of the centrality measures of the top 50 nodes*