Computational Study of Targets in Lung Adenocarcinoma

Grant Proposal

Avanti Moghe

Massachusetts Academy of Math and Science

Worcester, Massachusetts

Author Note

I would like to thank Dr. Crowthers for his guidance and assistance for me to be able to further my research.

Abstract (RQ) or Executive Summary (Eng)

The overall aim of this project is to analyze mutated protein structures in lung adecarcinoma in non-small cell lung cancers. Then the analysis data would be used to identify a drug capable of blocking the mutation thus reducing the susceptibility of the cancer cell towards drug treatments and potentially improving cancer success therapy rates in patients with non-small cell lung cancer. The results of this study should show an analysis of lung adecarcinoma, their gene expression profiles, and any counteracting drugs or peptide sequences against the cell.

Keywords: lung adecarcinoma, computer modeling,

Non-Small Cell Lung Cancer

Non-small cell lung cancer (NSCLC) is one of the most frequent cases of lung cancer and is attributed to nearly 85% of lung cancer worldwide, and over 25% of cancer-related mortalities worldwide. Being one of two main types of lung cancer, it is a prevalent issue that claims many lives around the world. It is the leading cause of cancer related mortalities in men, as well the second highest leading cause in cancer related mortalities for women. The number of NSCLC-related diagnoses and related mortalities is expected to continually keep increasing amongst different demographics, thus making it a prevalent issue that needs to be addressed. (Hendriks et al, 2024).

Non-small cell lung cancer is split into multiple subtypes, some of which include adenocarcinoma, squamous cell carcinoma, etc. Of this, the majority of the NSCLC cases are adecarcinoma cases, making it a widely studied subject. This includes a study of the cancer 's molecular mechanisms which contribute to tumorigenesis (Venugopal et al., 2019). Current forms of treatment for NSCLC include surgery, radiation therapy, and drugs such as chemotherapy, immunotherapy, or targeted agents. (Lok et al., 2023).

Lung Adenocarcinoma and Its Treatments

NSCLC is typically categorized into multiple subtypes, of which the more common one is adenocarcinoma (the focus of this study). Lung adenocarcinoma (LUAD) is a malignant tumor that originates from the glandular epithelium of the bronchial mucosa. LUAD typically occurs in the peripheral regions of the lungs, which covers the outer portions of the lungs away from the center. Due to the high concentration of blood vessels in the tissue that contain cancerous cells, it causes hematogenous metastasis, which is the rapid spread of cancer through the bloodstream. Therapeutic treatments for this type of cancer include radiotherapy, chemotherapy, targeted agents, and immunotherapy (Lu et al., 2025). Although common cause for adecarcinoma is exposure to tobacco smoke, it can also occur due to genetic and environmental factors, which makes up nearly 15-20% of adecarcinoma cases. Because of usually late diagnosis and resistance to treatments, the survival rate of patients with this tumor is relatively low, making it an important health concern. Difficulties with adenocarcinoma treatment are mainly due to the tumor heterogeneity within the said cancer, making it increasingly important to study cellular and molecular functions of the cells as well as the tumor microenvironment.

As of current studies for the tumorigenesis in lung adenocarcinoma, prevalent gene alterations include *KRAS, TP53, EFGR*, etc. These are specific location in the gene sequences that initiate the cancer within the cell and are thus extensively studied for targeted therapies at early stages of lung adenocarcinoma. Of these, the mutations in the *KRAS* oncogene are one of the more prevalent ones, with a poor prognosis and a particularly aggressive tumor. Various mutations in the epithelial cells can also lead to different outcomes as well as the tumor microenvironment, which also has an impact on the cells. The cell of origin for lung adecarcinoma has not yet been discovered, thus making it an important field of study (Seguin et al., 2022).

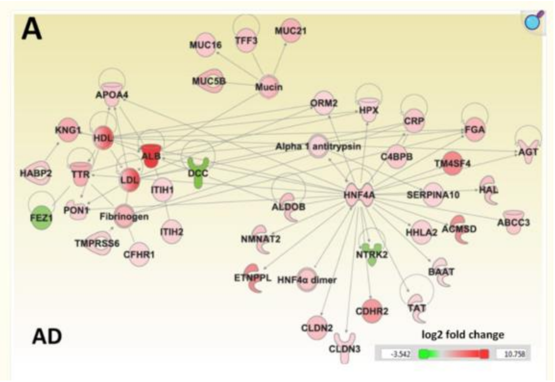

Fig 1 – gene/protein interaction in lung adenocarcinoma. From a paper by Venugopal et al.

Figure from paper by Venugopal et al, which shows interactions between networks of genes in lung adenocarcinoma.

Computational Analysis Methods

Through the modes of computational analysis, it is possible to closely study cancers such as lung adenocarcinoma at a higher level of detail and complexity, allowing researchers of new methods to battle these cancers. Computational models, genome datasets, and other tools make it possible for further research in understanding and subsequently. On the side is a flowchart from a paper by Khan et al., which details a simple way on how models are made to analyze biological processes.

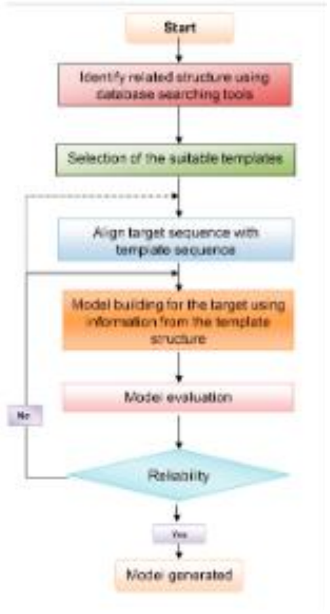

Fig 2 – Basic Flowchart of computation modeling

Current computational methods include Homology methods, *In-silico* methods, molecular dynamic simulations which are useful in protein and drug modeling and studying their formations and interactions with other molecules.

Homology modeling is the usage of current databases to construct a 3D replica of a protein that has an unknown experimental structure. In homology methods, the template, or the protein used from the reliable database that is used for creating the replica, must have a sequence similar to the target protein. This template sets the structure for the building of the main protein, which would then potentially be used for further analysis elsewhere.

In-silico methods are methods used for analyzing biological and physiological processes, making it possible to study and predict outcomes without actual experiments. These are quite useful for generating data and making predictions on various processes, mutations of cells, and other biological factors.

Molecular dynamic simulations are tools used for computationally studying biomolecules. This is extremely useful in studying proteins at a molecular level and thus discovering protein fluctuations and conformation. It allows for studying the interactions of the molecules and atoms, making it possible to closely observe the interactions of the proteins and ligands (Khan et al., 2016).

Various types of analyses are also capable of computation modeling, thus making it possible to analyze gene and protein interactions. The analyze used existing model and data to make conclusions about potential biomarkers, targeted therapies and other information about gene and protein interactions within cancer cells. Something like this was done in a study by Tomita et al., where the researchers built deep-learning models to analyze FFPE slides or Formalin-Fixed Paraffin-Embedded slides, which are slides with embedded with biological tissue, to find common genome mutations. Such models are validated when they exhibit a certain AUC or area under the curve, suggesting the accuracy of the said model. With this information on gene mutations, there can be a greater focus on precision medicine for cancers based on knowledge of gene level occurrences in lung adenocarcinoma. Such as in a study by Ruan et al, where the researchers used multiomics data to look at gene expression, DNA methylation, gene mutations, copy number variation data, and clinical data of LUAD patients, after which the research was used to subdivide the data into four clusters. These clusters a different type of common molecular subtype which was then used to analyze its reactions to prognosis, cancer therapies, and other factors important in cancer research.

Public databases are also used throughout this project (such as the cancer genome atlas, protein databank, GeneOmnibus) for the purposes of finding templates for similar proteins structure, analyzing mutations, ligand docking, and for other purposes of this research.

Section II: Specific Aims

This proposal's objective is to present my research objectives and aims for studying lung adenocarcinoma cells and the addition of drug sequences to mitigate the mutations in lung adenocarcinoma.

The long-term goal is to analyze gene profiles of lung adenocarcinoma using existing datasets from the cancer genome atlas, GeneOmnibus, and other publicly available datasets, and then subsequently using this data after analysis to observe its impact on tumorigenesis. After which the data will be used to identify the mutations which cause oncogenesis and cause the normal cells to become malignant. This would also include studying certain mutated proteins within the cell using online tools for observing and analyzing the protein. This knowledge would be subsequently used for looking for a drug or peptide sequence capable of blocking the normal cell from mutating and causing the tumor to further progress.

The rationale is that by researching and identifying the peptide sequence, that could potentially destabilize the malignant cell from furthering cell growth and division. Rather than conventional methods of chemotherapy, there could be a greater study on more targeted agents that can be used to combat cancer.

Specific Aim 1: Identify a prevalent gene mutation within the lung cancer cell that when deactivated or activated can cause the mutated cell to progressively become more malignant. Study mutation frequencies, which are more common for cell mutations, development from a normal cell to malignant cell, etc.

Specific Aim 2: Identify certain gene sequences or proteins and use this to look at certain mutated proteins which cause oncogenesis in lung adenocarcinoma. Analyze gene and protein interaction which will be used to get overall picture of the cancer model.

Specific Aim 3: Test simplified lung adenocarcinoma model with targeted drugs or peptide sequences to test whether the peptide can prevent the cell from mutating.

The expected outcome of this work is to identify a peptide sequence or drug capable of blocking the mutation of the cell from occurring.

Section III: Project Goals and Methodology

Relevance/Significance

Lung adecarcinoma accounts for a large percentage of cancer-related mortalities in the world. Despite multiple forms of treatment, it remains a very aggressive cancer with very low levels of survival rates for the patients. With this study, it may be possible to understand the mutations that cause this cancer's oncogenesis and therefore look more deeply into how this cancer can be treated in more effective methods.

Innovation

Computational modeling is a capable method to collect data on and analyze with high accuracy the biological and physiological processes. By harnessing the capabilities of these techniques, it would open new avenues to further in-depth research on specific targeted therapies for lung adenocarcinoma and potentially other cancers. Homology methods, *in-silico* methods, and MD simulations are all potential techniques to analyze complex level cells and make highly accurate predictions before moving on to in vivo or vitro testing.

Methodology

Examining Gene Profiles and Mutations

The beginning stage of this study is to determine the gene expression profiles of lung adenocarcinoma cells using public datasets such as the cancer genome atlas and Gene Omnibus. For this stage I will be collecting data from these public datasets on gene profiles in lung adenocarcinoma as well as normal lung tissue cells. These datasets will be used to compare against each other to identify overly expressed (or under expressed) genes within the mutated cells. This data will be compared to a normal lung tissue cell in order to find DEGs or differentially mutated genes, thus identifying more strongly expressed genes that contribute to oncogenesis. By obtaining this data, I will be able to observe and spot the mutation which causes the oncogenesis, or uncontrollable cell growth (Lu et al., 2025). Somatic mutations common in LUAD, so these will be the types of mutations I will investigate at depth and how the correlate with development of LUAD in the patient. This would include using existing models capable of taking in data and then using that information to track down most common gene mutations and Information on this mutation and its location will help me in narrowing down the specific protein I want to focus on for my project and analyze its pathways and interactions in the mutated cell.

Justification and Feasibility.

By obtaining data on gene profiles and mutations in malignant cells, I will be able to narrow down the specific gene sequence or protein I would like to focus on. I could obtain reliable datasets from public datasets such as the cancer genome atlas or Gene Omnibus which will be used analysis to study lung adenocarcinoma. With this information and by successfully completing specific aim 1 I will be able to proceed to my next steps which would include tracking my gene sequence to the specific mutated protein that promotes the oncogenesis within my tumor of focus, which is lung adenocarcinoma. Information on the mutated protein which causes tumor growth will assist me in building a simplified cancer model that will be used to analyze the development of the cancer cell as well as the influence of the protein on the development. Platforms such as Pymol and the Protein Data Bank will be essential in analyzing my protein and in making observations. My focus will then turn to research a counteracting drug or blocking peptide sequence for the the lung adenocarcinoma. Computational methods will be used to analyze how well the drug or peptide blocks the development caused by certain mutated proteins in cancer cells. This study would be done in the early stages of the cancer as heterogeneous metastasis in LUAD tumors cause the cancer to spread quite from original locations so it would be difficult to implement the drug or peptide in such situations. So my focus would be on specifically early stages of the cancer on how the drug or peptide can work in concentrated conditions, and future research would include then scaling the drug to see if it can work in different stages.

Summary of Preliminary Data. Using GEPIA and cBioPortal which are webtools, I conducted a differential gene expression analysis and mutation frequency analysis. These analyses determined that lung adenocarcinoma was driven by somatic mutations and that these high number of mutations were most likely due to copying extra of certain chromosomes during DNA replication. More specifically I looked at mutations that impacted the KRAS and EGFR proteins such as

Figure 1: Pacific sand lance (*Ammodytes hexapterus*) burrowing into the sand. Mandy Lindeberg, NOAA/NMFS/AKFSC - http://www.photolib.noaa.gov/htmls/fish1917.htm

Expected Outcomes. The overall outcome of this aim is to further research on cancer development and how current technologies can be used to deepen the study on more targeted therapies for cancer and perhaps other diseases. This knowledge will be used for…..

Potential Pitfalls and Alternative Strategies. We expect….

Specific Aim #2:

Please follow the section format described above for an individual-specific aim.

Section III: Resources/Equipment

Section V: Ethical Considerations

Section VI: Timeline

Section VII: Appendix

Section VIII: References

Hendricks, L. E. L., Remon, J., Faivre-Finn, C., Garassino, M. C., Heymach, J. V., Kerr, K. M., Tan, D. S. W., Veronesi, G., & Reck, M. (2025). Non-small-cell lung cancer. *Nature*, *10*. https://doi.org/https://doi.org/10.1038/s41572-024-00551-9

Seguin, L., Durandy, M., & Feral, C. C. (2022). Lung adenocarcinoma tumor origin: A guide for personalized medicine. *Cancers*, *14*(7), 1759. https://doi.org/10.3390/cancers14071759

Khan, F. I., Wei, D.-Q., Gu, K.-R., Hassan, Md. I., & Tabrez, S. (2016). Current updates on computer aided protein modeling and designing. *International Journal of Biological Macromolecules*, *85*, 48–62. https://doi.org/10.1016/j.ijbiomac.2015.12.072

Venugopal, N., Yeh, J., Kodeboyina, S., Lee, T., Sharma, S., Patel, N., & Sharma, A. (2019). Differences in the early stage gene expression profiles of lung adenocarcinoma and lung squamous cell carcinoma. *Oncology Letters*. https://doi.org/10.3892/ol.2019.11013

Lok, E., Liang, O., Malik, T., & Wong, E. T. (2023). Computational analysis of tumor treating fields for non-small cell lung cancer in full thoracic models. *Advances in Radiation Oncology*, *8*(4), 101203. https://doi.org/10.1016/j.adro.2023.101203

Lu, Z., Bao, P., Wang, T., Hu, K., Zhang, L., Yi, L., Pan, Y., Li, W., Lu, Z. J., Wang, J., & Ruan, J. (2025). Multi-omics data-based modeling reveals tumorigenesis- and prognosis-associated genes with clinical Tomita, N., Tafe, L. J., Suriawinata, A. A., Tsongalis, G. J., Nasir-Moin, M., Dragnev, K., & Hassanpour, S. (2022). *Predicting Oncogene Mutations of Lung Cancer Using Deep Learning and Histopathologic Features on Whole-Slide Images*. https://doi.org/10.1101/2022.05.03.22274614 potential in lung adenocarcinoma. *BMC Cancer*, *25*(1). https://doi.org/10.1186/s12885-025-14943-x

Tomita, N., Tafe, L. J., Suriawinata, A. A., Tsongalis, G. J., Nasir-Moin, M., Dragnev, K., & Hassanpour, S. (2022). *Predicting Oncogene Mutations of Lung Cancer Using Deep Learning and Histopathologic Features on Whole-Slide Images*. https://doi.org/10.1101/2022.05.03.22274614

Ruan, X., Ye, Y., Cheng, W., Xu, L., Huang, M., Chen, Y., Zhu, J., Lu, X., & Yan, F. (2022). Multi-omics integrative analysis of Lung Adenocarcinoma: An in silico profiling for precise medicine. *Frontiers in Medicine*, *9*. https://doi.org/10.3389/fmed.2022.894338