**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).

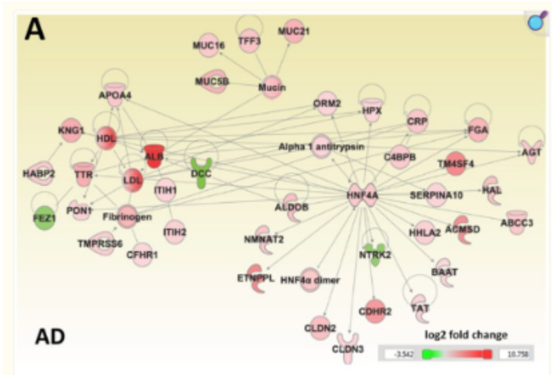


Fig 1 – gene/protein interaction in lung adenocarcinoma. From a paper by Venugopal et al.

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.