Blog post VICOM – LUCIA January 2024.
Deep Phenotyping for Lung Cancer through Artificial Intelligence: The LUCIA project

Authors: Alba Garin-Muga, Iván Macía (VICOMTECH)

Precision medicine, often termed personalized medicine, aims to provide a more efficient approach to prevention, diagnosis and treatment by analyzing characteristics through analysis and modelling of large volumes of genotype or phenotype-rich data (demographics and lifestyle information, clinical data in the electronic health records, biomarkers, genomics, medical imaging…). Individuals may be also divided into groups based on these differential characteristics or patterns, a process which is known as patient stratification, so they can be treated similarly. The implementation of precision medicine requires new approaches for the systematic and precise analysis of all the wealth of biomedical information available, such as deep phenotyping.

Deep phenotyping is defined as the precise and comprehensive analysis and characterization of phenotypic traits and abnormalities in which the individual components of the phenotype are observed and described, often for the purposes of scientific examination of human disease [1]. To achieve one of the main goals of personalised medicine, the stratification of patients into groups with a common biological basis of disease, there is a need for an in-depth analysis that involves collecting a wide range of data, including clinical information from EHR, radiological imaging, genomic and other omics sequencing, and tissue imaging data from pathology samples describing the tumour microenvironment and other relevant tissue, cellular or molecular features.

Lung cancer is a heterogeneous disease, driven by accumulated genomic mutations, that varies widely among individuals and benefits deeply from precision medicine approaches. Through deep phenotyping, clinicians can gain a more precise understanding of the biological insights associated with each patient's cancer. The vast amount of complex data generated for deep phenotyping makes this scenario suitable for the use of Artificial Intelligence (AI) for modeling. AI technologies, particularly machine learning (ML) and deep learning (DL) offer several advantages to efficiently analyze, model, and extract meaningful patterns and insights from large, complex datasets related to phenotypic information, suitable for explaining differences in patient outcomes or treatment resistance. Additionally, incorporating diverse data modalities offers prospects to enhance the resilience and precision of diagnostic and prognostic models, bridging the gap between AI and clinical practice.

In the LUCIA project, we intend to examine retrospective clinical cases and diagnostic methods (such as lung CT, pathology, and omics) to investigate lung cancer phenotypes. This will be achieved through complex data integration utilizing AI, with the aim of gaining insights into the underlying biology and its correlation with the genome and the risk factors. Applying deep phenotyping in lung cancer involves a multidisciplinary approach. Therefore, in the LUCIA project, partners from across Europe specializing in lung cancer research, diagnostic technologies, artificial intelligence, ethics and legal frameworks, and EU policy, along with four hospitals, will work together to integrate various technologies to comprehensively assess and characterize the disease. During the project, several key steps will be applied:

- Patient data collection. In the project, we will gather lung cancer data in the LUCIA health data platform. The health data platform will enable lung cancer data collection, storage, management, preparation and analysis. It will allow the storage of both retrospective and prospective data, such as data from electronic health records, lifestyle or exposure information collected from questionnaires, registries, or open data sources, and store them in a virtual data lake, where the data is accessible through a standard interface for use in data research studies and AI model development.
- Data processing and analysis: The LUCIA health data platform will consolidate extensive data from both existing data sources and generated within the LUCIA project. Within the project, the data will be ingested, structured and cleaned, stored, analysed and visualised using data analytics and AI techniques. AI algorithms play a crucial role in swiftly and precisely processing, organizing, and analyzing this data. Machine learning and deep learning techniques can uncover patterns, relationships, and anomalies within the data that might not be readily apparent to human researchers. Electronic health records, lifestyle and geo-location information, non-invasive device information, CT scans, pathology slides, omics, … data, due to their complexity, require advanced techniques for their analysis and interpretation.
- Data integration and AI modeling: Within the project, all the gathered data will be combined to create a comprehensive patient profile. Also, ML and DL models will be built to characterize the disease, detect biomarkers and predict the probability of suffering from lung cancer based on the whole patient information.
- Data visualization: The virtual research environment that is being developed within the project will enable researchers and analysts to derive insights, patterns, and trends from health-related data. It should provide powerful tools for data exploration, analysis, and presentation.

All these steps will allow an exhaustive examination of the biological mechanisms underlying the several disease phenotypes in order to understand existing biomarkers for more precise diagnosis and to predict the outcome or response to treatment.

To sum up, deep phenotyping gathers details on disease manifestations in a more individualized and detailed manner, employing advanced algorithms to seamlessly integrate the resulting wealth of data with various types of information. As Isaac Kohane, renowned for his contributions as a bioinformatician at Harvard Medical School in Boston, Massachusetts, states, deep phenotyping “shows the different dimensions of the disease.” The LUCIA project aims, through a multidisciplinary approach, to contribute to the understanding of the disease, the identification of biomarkers, early detection, and improved management of the illness, ultimately saving patient lives.

References

[1] Robinson, P. N. (2012). Deep phenotyping for precision medicine. *Human mutation*, *33*(5), 777-780.

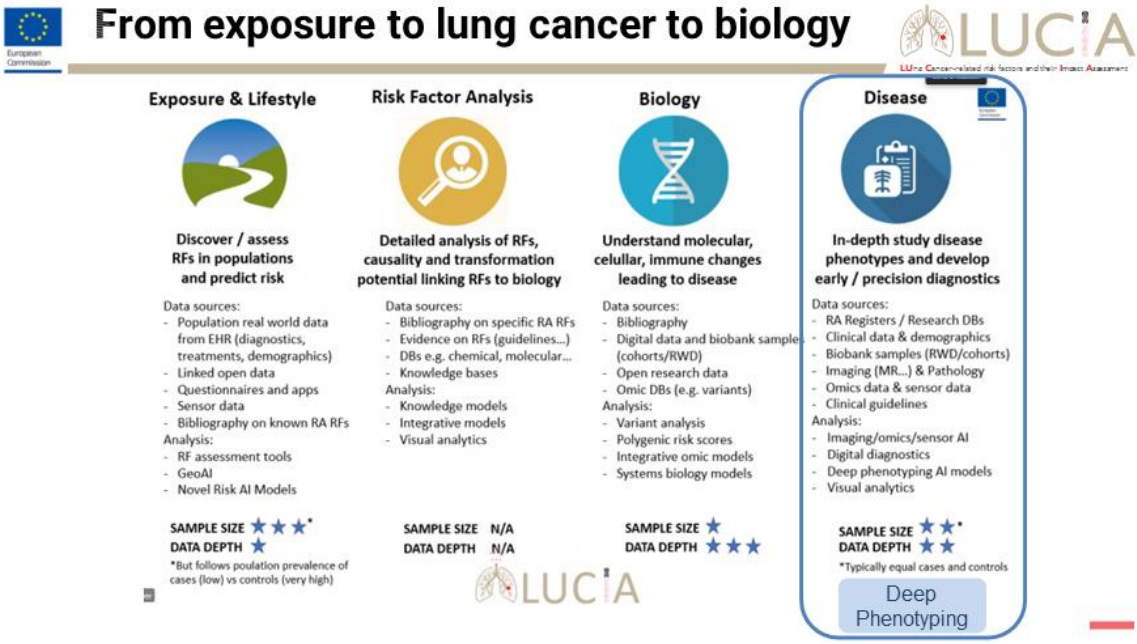

Figure 1: Clinical studies, data sources, and their objectives. Deep phenotyping requires a medium-to-large sample size and data depth to unravel the biological insights of the disease that enable the early diagnosis of the patients.

The LUCIA project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement no. 101096473. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.