Identification of sub-groups of chronic obstructive pulmonary disease patients using clustering analysis of small airway histopathology

Background

Chronic obstructive pulmonary disease (COPD) is a debilitating lung condition characterised by poorly reversible and often progressive airflow obstruction. COPD is caused by the inhalation of noxious particles and gases, most commonly cigarette smoke. The small airways, defined as <2mm in diameter, are the major anatomical location of increased airflow resistance in COPD. Pathophysiological changes in the small airways such as epithelial or airway wall thickening can narrow the lumen, while reduced alveolar tethering support to the small airways can cause partial or complete collapse during expiration. However, these changes are highly heterogeneous and it is unclear if sub-groups of patients with distinct features of small airway remodelling exist. Identifying sub-groups of COPD patients is essential for understanding shared molecular mechanisms and the development of tailored treatments.

Aim

The aim of this study is to cluster COPD patients using histopathological measurements of small airway remodelling and evaluate how this relates to the clinical demographics of disease.

Methods

Using a biobank of human distal lung tissue, small airway histopathological measurements, including epithelial thickness, wall thickness and alveolar attachment destruction have been collected from >200 COPD patients and >100 controls. Using this data, COPD patients and controls will be clustered. A range of clustering techniques will be used. We will also seek to explore the potential of this biobank for future research and produce a report

HPC and software engineer

High performance computing (HPC) and software engineering are essential for this project because of the scale and complexity of the data. Analysing histopathological measurements from hundreds of patients, combined with clinical variables, requires substantial computational power for efficient data processing, clustering, and statistical modelling. HPC enables these analyses to be performed quickly and at scale. Strong software engineering practices ensure the analysis pipeline is reproducible, well-structured, and scalable, allowing models to be reliably applied, updated, and shared, which is critical for robust scientific findings and future clinical translation.