Quality over quantity: the importance of collecting relevant samples to understand complex diseases

Alen Faiz¹,²,³ and Maarten van den Berge²,³

¹University of Technology Sydney, Respiratory Bioinformatics and Molecular Biology (RBMB), School of Life Sciences, Sydney, Australia. ²University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD (GRIAC), Groningen, The Netherlands. ³University of Groningen, University Medical Center Groningen, Dept of Pulmonary Diseases, Groningen, The Netherlands.

When investigating the mechanisms behind heterozygous diseases such as COPD, collecting and sequencing samples and comparing it back to histology can help find new insights into the development of disease. https://bit.ly/3CV49O1

Cite this article as: Faiz A, van den Berge M. Quality over quantity: the importance of collecting relevant samples to understand complex diseases. Eur Respir J 2022; 59: 2200418 [DOI: 10.1183/13993003.00418-2022].

Whole transcriptional studies investigating COPD have provided important insights into the development and progression of this complex disease. These have ranged from transcriptome studies on easily accessible samples like blood, induced sputum and nasal epithelial brushes [1], to those that are more invasive and difficult to collect, such as bronchial biopsies and brushes [2, 3], or lung tissue [4, 5], which usually requires tissue derived from transplant donors or lung cancer resection. Lung tissue is considered one of the most relevant samples to COPD due to the nature and location of parenchymal tissue destruction. However, the heterogeneity of the disease, not only between patients but also within a single individual, makes interpretation of the results more difficult. In the article by Feng et al. [6] in the current issue of the European Respiratory Journal, the authors show that the loss of terminal bronchioles in COPD occurs in so-called “hot spots”, which are areas with a linear mean intercept between 500 and 1000 µm. Transcriptional profiling of lung tissue from these hot spots using “regional transcriptomics” identified an 11 gene signature associated with upregulation of IFNG signalling, co-stimulatory immune checkpoint genes, and genes related to the inflammasome pathway. Regional transcriptomics is a method where multiple samples are taken from the same patient, and matching histology is collected from samples that are sequenced. This method allows for histological features to be linked to transcriptomic profiles and potentially helps to identify the genes and mechanism behind these features (figure 1a).

Previous large studies comparing COPD lung tissue to non-COPD lung tissue have not found the 11 genes to be significant between the two groups or COPD disease severity [5]. A possible explanation for the fact that this signature was not found to reflect COPD or COPD disease severity may be the lack of data on regional emphysema severity. The data provided by Feng et al. [6] indicate that this information should be taken into account in future studies collecting lung tissue samples. Identifying gene expression changes associated with regional emphysema severity is not a new concept, with Campbell et al. [7] identifying 127 genes that were significantly associated with regional emphysema severity. This study differed from Feng et al. [6] as it focused solely on the degree of emphysema and not on the loss of terminal bronchioles. As could be expected, no overlap between the signatures derived from the two studies was observed. One thing to note is that regional transcriptional studies tend to suffer from power issues as they usually require multiple samples from each patient, with Campbell et al. [7] only having two healthy individuals and six patients with COPD.

Feng et al. [6] also used cellular deconvolution through a program known as CIBERSORT [8], to determine the inflammatory cell infiltration into the “hot spots”. This method determines the percentages of certain cell types in a given sample based on the expression of marker genes unique to the cell types being tested, and has been previously used to determine the composition of bronchial biopsies [9]. Using this...
method, they found that CD4, CD8, and B cell lymphocytes were infiltrating the hotspots. This nicely links to a previous report that found the severity of emphysema measured by parametric response mapping was associated with higher levels of CXCL11 expression in the airways, which is a potent chemokine involved in CD8+ T cell activation [10]. To verify their findings, FENG et al. [6] used serial sections from the same region that they sequenced, so that they were able to verify gene expression changes in inflammatory cells by immunohistochemistry. Cellular deconvolution together with regional transcriptomics is a powerful combination to identify and verify changes in cellular composition in a region-based analysis.

Single-cell sequencing (sc-Seq) analysis provides a profile of the transcriptome for each individual cell and offers another method to overcome the heterogeneity of lung tissue samples. Sc-Seq studies of COPD parenchyma are limited but have identified a subpopulation of alveolar epithelial type II cells that based on transcriptomic data was predicted to have reduced cellular stress tolerance in COPD [11]. This finding would not have been elucidated with bulk RNA-Seq, which carries a large amount of noise due to differences in cellular composition. In the context of the findings of FENG et al. [6], this reduced cellular stress tolerance may extend to the terminal bronchioles. Interestingly, taking the 11 gene signature and applying it to sc-Seq data from parenchyma (n=6) [12], we see that this signature is mainly expressed in neutrophils, dendritic cells and luminal macrophages (figure 1b). The latter nicely matches their finding of increased M1 macrophages in their “hot spot” region. A drawback of sc-Seq is that it involves dissociation

**FIGURE 1** Overview of the methods to measure spatial information in tissue samples. a) Types of RNA sequencing and their ability to retain spatial information. b) U-map of parenchyma samples (n=6) displaying the expression of the 11 gene “hot spot” signature.
of cells during sample processing; therefore, linking the cells back to their original locations has been challenging. Interestingly, recent sequencing advances have led to a method to overcome this dissociation issue. Spatial transcriptomics is a method of performing RNA sequencing directly on tissue sections, allowing the cellular structure to be linked to transcriptomic profiles. This technology has yet to be used in COPD; however, it has been suggested that it will move the field forward and has been nicely explained in a recent review by Curras-Alonso et al. [13].

One of the main findings of Feng et al. [6] is that terminal bronchioles are destroyed within regions of microscopic emphysematous destruction with an average airspace size of \(\geq 500 \text{ and } <1000 \mu\text{m}\) in COPD. This provides some molecular mechanism behind the observation that the destruction of the terminal bronchioles precedes the onset of emphysematous destruction measured by micro-computed tomography and the destruction of the terminal bronchioles [14].

In conclusion, spatial and regional transcriptional studies provide a promising new tool for developing clinically relevant biomarkers for the development of emphysema and provide novel treatment targets that could lead to improved personalised treatment of this disease. Feng et al. [6] provide one of the first well-powered studies to investigate this transcriptional spatial information.

Conflict of interest: All authors have nothing to disclose.

References

https://doi.org/10.1183/13993003.00418-2022