Another step in COPD-endotyping
Ditz, Benedikt

DOI:
10.33612/diss.245078633

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 1.

General introduction
1.1 COPD

Chronic Obstructive Pulmonary Disease (COPD) is one of the leading causes of morbidity and mortality worldwide\(^1\). COPD is characterized by respiratory symptoms (i.e. shortness of breath, cough, and/or sputum production) and incompletely reversible airflow obstruction\(^1\). A ratio of $< 70\%$ of the post-bronchodilator forced expiratory volume in one second ($\text{FEV}_1$) to the forced vital capacity ($\text{FVC}$) defines COPD according to the guidelines of the "Global Initiative for Chronic Obstructive Lung Disease (GOLD)\(^1\). The chronic airflow obstruction is caused by a mixture of small airways disease and parenchymal destruction (emphysema) and becomes clinically apparent around 40-50 years of age. Chronic inflammatory processes are leading to a tissue-repair and-remodeling process, that increases mucus production in the airway lumen, enlarges the bronchial mucus glands, increases the proliferative activity of airway epithelial cells, and increases both mucous and squamous cell metaplasia\(^2\). Tobacco smoking and occupational inhalant exposures are considered to be the primary risk factors, however, there are additional risk factors in very early life that likely drive the development of COPD in later life\(^1\)-\(^3\). In utero and childhood lung development and exposures, such as maternal smoking, but also recurrent respiratory tract infections and childhood asthma represent relevant early life risk factors that can drive COPD development (Figure 1).

COPD represents an important public health challenge that is both preventable and treatable\(^1\). However, appropriate treatment decisions in COPD patients still represent a challenge for clinicians, since COPD exhibits various clinical characteristics. Although the chronic airflow limitation represents the key characteristic of COPD, the relative contribution of small airways disease and parenchymal destruction differs from patient to patient and varies within the same patient over time\(^3\). In addition, COPD-related chronic inflammatory processes do not follow a strict pattern in affected patients, but vary between patients and evolve at different rates over time. Consequently, the view on COPD is changing from an airflow limitation-centric view to the realization that COPD is a complex and heterogenous disease\(^1\)-\(^4\). It has become clear that assessing COPD solely by lung function variables is insufficient to address its heterogeneity and that clinicians require additional tools that help them to better differentiate patients and to provide therapeutic interventions in a personalized way\(^4\). The development of such tools represents a major focus in COPD research over the last two decades. One popular approach represents the concept of “COPD phenotyping”, which was introduced by Han et al. in 2010, which considers “a single or combination of disease attributes that describe
differences between individuals with COPD as they relate to clinically meaningful outcomes (symptoms, exacerbations, response to therapy, rate of disease progression, or death)\(^5\). Classification approaches of phenotypes in COPD were already made more than 30 years ago by distinguishing between two types of COPD patients, one with chronic bronchitis ("blue bloater": the presence of productive cough more than 3 months per year in two or more consecutive years) and the other with emphysema ("pink puffer")\(^6,7\). Additional phenotypes have been suggested over the past years, for example, the frequent exacerbator (presence of two or more exacerbations per year), asthma-COPD overlap (persistent airflow limitation in (ex-)smokers with several asthma-associated features), the fast decliner phenotype (a rapid decline of lung function) or the comorbidities and systemic phenotype (high comorbidities burden, predominantly cardiovascular and metabolic)\(^7,8\). Nevertheless, it is unlikely that addressing the complexity and heterogeneity of COPD solely by clinical phenotyping is sufficient to provide therapeutic interventions in a personalized way\(^4,9\). Different phenotypes often occur simultaneously in the same patient and vary in their contribution to the individual disease burden.

![Figure 1: Graphic representation of the risk factors for chronic obstructive pulmonary disease during the different stages of life. Reprinted with permission of Lancet. Copyright © 2015 Elsevier. Postma DS et al. Lancet 2015; 385:899-909.](image-url)
Identifying “phenotypic traits” is likely a necessary intermediate component towards personalized medicine in COPD\cite{4}. The concept of clinical phenotyping is evolving to the paradigm of “treatable traits”, which represents a “label-free” management strategy based on “treatable traits” in each patient\cite{10}. “Treatable traits” refer to patient characteristics that can be targeted for therapeutic intervention. In principle, these traits can be identified by clinical, histological, radiographic, or biological markers. For example, COPD patients with emphysema and without interlobar collateral ventilation (a phenotypic trait identified by imaging) can be targeted for endobronchial valve placement\cite{11}.

The concept of “treatable traits” is not only based on clinical phenotyping but also aims to address the biologic heterogeneity in the form of “endotypes”, in preparation for targeted therapy. Endotyping represents an emerging dimension in COPD research, aiming to assess and stratify molecular pathobiological mechanisms related to the disease (Figure 2)\cite{12}. To identify endotypes in COPD, it is necessary to explore the biological heterogeneity. Therefore, there is an increasing interest to implement technologies that allow assessing patient (host) genetics as well as interactions between the host and cumulative environmental exposures (e.g. the microbiome, smoking). High-throughput mass screening technologies, known as “-omics”, represent promising approaches to address this request, by enabling system-wide profiling of the genome, transcriptome, epigenome, microbiome, or metabolome\cite{13}. Currently, the best example of a well-recognized endotype in COPD is alpha-1 antitrypsin deficiency, which is associated with panacinar emphysema and has validated biomarkers at gene and protein level\cite{14}. Further emerging endotypes in COPD include the assessment of airway inflammatory profiles, such as T-helper cell type-driven inflammation endotypes\cite{12,15,16}. T cells play a crucial role in COPD, and different subsets are involved in orchestrating inflammatory processes which result in distinct inflammatory and structural consequences\cite{15}. For example, type 17 helper T cells (T-17) are associated with increased airway obstruction, functional small airways disease and decreased response to corticosteroids, irrespective of airway eosinophilic inflammation and mechanistic data suggest that T-17 may play an important role in the pathogenesis of emphysema\cite{16,17}. Another example represents type 2 inflammation (T-2), which is mediated by helper T type 2 cells and innate lymphoid, basophils, eosinophils, and mast cells\cite{12}. While marking a key endotype of asthma, and exhibiting characteristics such as airway hyperresponsiveness, airway and blood eosinophilia, and preferential response to ICS, T-2 may represent a distinct endotype as well in COPD\cite{12,18,19}. Christenson and colleagues have shown that a genomic signature of T-2 inflammation was associated with increased airway biopsy
and blood eosinophils, greater bronchodilator responsiveness, and improvement in hyperinflation with ICS treatment in COPD subjects without a history of asthma\textsuperscript{19}. Overall, the concept of endotyping exhibits a promising potential to get more insight into the biological and clinical complexity of COPD. The development of endotype-specific biomarkers that can address and stratify its biological heterogeneity would represent a major improvement regarding the precision of diagnosis as well as facilitate treatment decisions of existing and new therapies.


1.2 BASIC PRINCIPLES OF TRANSCRIPTOMICS

Almost every human cell contains the same genome, however, not every gene is transcriptionally active in every cell. Transcriptomics involves the sequencing of the entire transcriptome within a given sample, focusing on the quantitative and qualitative characterization of ribonucleic acid (RNA) transcripts across a genome.
Only a small percentage (<5%) of the human genome is transcribed into RNA molecules. The transcriptome reflects the dynamic processes (gene expression) in a given tissue and consists of coding (messenger: 1-4%) and non-coding (ribosomal, transfer, small nuclear, small interfering, micro, and long-non-coding: >95%) RNA while capturing a snapshot in time of all transcripts present in a cell. Understanding the transcriptome allows us to interpret the functional elements of the genome and explore disease-specific pathophysiological processes under given conditions. For example, Beane and colleagues have shown that pathways related to the metabolism of xenobiotics by cytochrome P450 and oxidoreductase activity were enriched among genes differentially expressed in airway epithelial cell brushings from smokers compared to never-smokers. This study represents one example, demonstrating the potential of transcriptomic sequencing providing new insights into the biology of the airway field of injury associated with smoking.

In the past, gene expression studies used to be restricted to quantify individual gene transcripts. For example, reverse transcriptase quantitative polymerase chain reaction (PCR) analyses allow measuring expression profiles of a limited number of predefined genes. Over the last two decades, the situation has dramatically changed due to the development of the two key contemporary technologies microarrays and RNA-sequencing (RNA-Seq). Both methods allow the profiling of tissue's genome-wide RNA transcripts based on different technological approaches, but both begin with (i) purifying high-quality RNA of interest and (ii) converting the RNA to complementary DNA (cDNA). For microarrays, the cDNA is fluorescently labeled and hybridized to probes on a chip, which is used to quantify RNA abundance. This method is relatively inexpensive, but it also has important limitations. Transcripts not included in the probes will not be detected. Therefore, a priori sequence knowledge is required about the organism's genome of interest to detect and quantify known transcripts, whereas the discovery of de novo transcripts will not be possible.

For RNA-Seq, RNA extracted is followed by mRNA enrichment or ribosomal RNA depletion and the cDNA is fragmented and converted to a library of cDNA fragments with adaptors attached to one or both ends. Then, these molecules are sequenced in a high-throughput manner from one end (single-end sequencing) or both ends (pair-end sequencing) generating short reads (30-400 bp). The next steps include a computational reconstruction of these reads by either de novo assembly without a genome (eg. Trinity methodology) or aligned to a reference genome/transcript. Finally, the quantified reads need to be filtered and normalized between samples to allow statistical modeling of significant changes of individual gene expression levels between sample groups.
It is fair to say that RNA-Seq has surpassed microarray sequencing since it exhibits several key advantages over microarray transcriptomes. First, RNA-Seq is not limited to a pre-defined set of probes but measures all transcripts present in a sample\textsuperscript{26}. Further, the amount of input RNA is much lower for RNA-Seq (nanogram quantity) compared to microarrays (microgram quantity) and allows a finer examination of cellular structures. In addition, RNA-Seq captures a large dynamic range of expression levels of transcripts, since it does not have an upper limit for quantification and performs better concerning the detection of low abundance transcripts\textsuperscript{26}. Therefore, expression profiles of genes and associated pathways can be studied in much higher detail by RNA-Seq compared to the microarray.

Figure 3: Overview of the types of variants in the genome, their potential consequences and the methods/techniques to untangle them. Reprinted with permission of Briefings in Bioinformatics. Copyright © 2018. Manzoni C et al. Briefings in Bioinformatics 2018; 19(2):286-302.
1.3 THE RESPIRATORY MICROBIOME IN COPD

The concept of the human microbiome was proposed by Joshua Lederberg in 2001, which is defined by the collective genomes of all microbiota inhabiting a specific environment/niche in the human body\(^2\). Traditionally, healthy lungs have been considered to be sterile, since standard microbiological culture techniques failed to grow microbes from airway samples in the past\(^2\). However, the establishment of culture-independent high-throughput sequencing techniques has led to a change of this paradigm, by showing that the lungs harbor a low-biomass of microbiota, both in healthy subjects and patients with respiratory diseases\(^3,4\). Therefore, the respiratory microbiome has gained increasing scientific attention over the past decade. There is an increasing body of scientific evidence, showing that the respiratory microbiome is associated with airway inflammatory processes, also in COPD, highlighting its potential role for COPD-endotyping (Figure 2)\(^5\,\,6\). Nevertheless, its role in COPD pathobiology and clinical utility, for example as a biomarker, is still in its infancy and requires standardization of methods, including sample collection, processing, and analytical pipelines.

1.4 OUTLINE OF THIS THESIS

This thesis aims to contribute to the field of endotyping in COPD by the use of transcriptomics across the host genome as well as the microbiome. Both exhibit a promising potential to further elucidate COPD-related pathology and for COPD-endotyping, which will be discussed in this thesis.

In chapter 2, we apply microarray and RNA-seq in bronchial biopsies of participants with COPD, before and after 6 months of inhaled corticosteroids (ICS) treatment. We investigated whether gene expression profiling, using RNA seq, provides additional insight into the transcriptional effects before and after ICS, compared to microarrays. Next to differential expression analysis, both RNA-seq and microarray datasets were analyzed for bulk tissue cell-type deconvolution, since the gene expression data were derived from a mixed cell population. The concept of cell-type deconvolution utilizes cell-type-specific gene expression from single-cell RNA sequencing data to characterize cell-type composition from bulk RNA-Seq or microarray data in complex tissues\(^7\). This is of interest since bulk transcriptomic data can result in a loss of signal from lowly abundant cell types and certain cell types are more vulnerable to disease than others\(^8,9\). For deconvolution, we first performed
AutoGeneS to select informative genes and used two different regression methods to infer cell-type proportions: nonnegative least squares (NNLS) and Support Vector Regression (SVR)\(^3\).

**Chapter 3** focuses to assess how smoking affects the transcriptome of the bronchial mucus barrier in COPD and healthy subjects. This study was based on a candidate-gene approach, including genes involved in airway surface hydration and mucus production. In addition, bulk RNA-seq cell-type deconvolution was performed to determine differences in cell-type composition. RNA-seq data was assessed in bronchial biopsies from three clinical studies including a COPD cohort, a healthy cohort, a stop-smoking cohort as well as a short-term mouse model of experimental COPD.

The positioning of ICS in the treatment of patients with COPD is under discussion since only a subset of COPD patients benefits clinically from using ICS, whereas they are ineffective or even associated with adverse effects in others\(^1,4\). Sputum and blood eosinophil levels have been identified as biomarkers for the efficacy of ICS in COPD, however, it can be argued whether these are sufficient to guide ICS positioning in COPD patients\(^1,4\). In **Chapter 4**, we performed whole-transcriptome profiling (RNA-Seq) in COPD sputum samples to investigate whether a sputum gene signature exhibits a higher predictive value for predicting COPD exacerbations after ICS withdrawal than sputum eosinophilia.

The sputum sample is an easily-obtainable and patient-friendly specimen, which makes it a preferred sampling technique for a potential airway biomarker in COPD. Determining the reproducibility and responsivity of cytokine measurements represents crucial milestones on the path of finding a robust sputum biomarker, which comes with several technical challenges. Some patients are not able to cough up spontaneous sputum and induced sputum samples can suffer from dilution effects by nebulized saline solutions, resulting in methodological problems regarding the determination of cytokines at the protein level. In **Chapter 5**, we investigated gene transcripts from several cytokine candidate-genes regarding their reproducibility and responsivity in induced sputum from COPD patients to assess their potential as a future sputum biomarker.

**Chapter 6** of this thesis discusses the current evidence, limitations, and future perspectives of sputum microbiome profiling in COPD. In addition, this review provides an overview of the basic principles of current high-throughput techniques for microbiome assessment and analytical approaches.
Transcriptomic profiling of the respiratory microbiome is still in its infancy, and current knowledge still relies mainly on other high-throughput technique approaches, especially 16S rRNA gene sequencing. The study in chapter 7 aimed to identify a biological signature of the viable microbiome in bronchial biopsies from smokers and ex-smokers with stable COPD, using an unbiased metatranscriptomic approach. Here, we aimed to compare the bronchial microbiome depending on smoking status as well as to study host-microbiome interactions.

Chapter 8 aims to summarise and discuss the findings of this thesis and seeks to issue future perspectives in the field of COPD-endotyping and transcriptomics of the host and respiratory microbiome.
REFERENCES


* shared first author
# shared last author

1University of Groningen, University Medical Center Groningen, Department of Pulmonary Diseases, Groningen, The Netherlands
2University of Groningen, University Medical Center Groningen, GRIAC (Groningen Research Institute for Asthma and COPD), Groningen, The Netherlands
3University of Groningen, University Medical Center Groningen, Department of Pathology & Medical Biology, section Medical Biology, Groningen, The Netherlands
4Institute of Computational Biology, Helmholtz Centre, Munich, Germany
5Technical University of Munich, Department of Mathematics, Munich, Germany
6Leiden University Medical Center, Department of Pulmonology, Leiden, the Netherlands
7OMNI Biomarker Development, Genentech Inc, South San Francisco, USA
8European Research Institute for the Biology of Ageing, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
9University of Technology Sydney, Faculty of Science, Ultimo NSW 2007, Australia