General discussion

This thesis aimed to identify novel risk factors for COPD, lower lung function, asthma and depression and to perform integrative studies to investigate the functional role and interaction of multiple omics layers. This chapter will emphasize the main findings, discuss challenges, possible clinical implications and future directions of COPD research.

FINDINGS OF THIS THESIS

COPD as a common and complex disease is a major public health burden.[1,2] The aim of this thesis has been to identify novel molecular determinants of COPD, lower lung function and related pathology such as depression and to perform integrative studies to investigate the functional role and interaction of multiple omics layers. GWAS identified common genetic variants to be associated with COPD, however, they exert small effects and their functional role is unknown. In **Chapter 2.1** of this thesis, I performed a genome-wide linkage scan to identify rare genetic variants associated with COPD. Genetic linkage analysis is a powerful tool to identify genomic regions shared among the affected family members. This was done in the large genetically isolated Erasmus Rucphen Family study. Using a genetic isolate, characterized with shared lifestyle and environment effects as well as less genetic variance, increases the power of the analysis. I found significant evidence for extensive linkage of COPD to the known COPD GWAS region, chromosome 15q14-15q25 and to two novel regions, 11p15.4-11q14.1 and 5q14.3-5q33.2. More importantly, I was able to identify four pathogenic, rare variants, shared by family members with COPD using exome-sequence data. They belong to the chromosome 11 peak, genes *AHNAK, PLCB3, SLC22A11* and *MTL5*. The variants in *SLC22A11* and *MTL5* were confirmed in association with COPD in our meta-analysis of 9,888 cases and 27,428 controls. These two genes are both expressed in lung tissue and are interesting candidates. Further functional studies should confirm their role in respiratory pathology in general population. Although I confirm the linkage of chromosome 15 region, I could not identify any shared rare variants by means of exome-sequence data. This may be explained by the fact that only common variants are relevant, or, alternatively that common or rare regulatory variants outside of exons may play a key role in COPD. Future candidate-gene studies using whole-genome sequencing should further investigate the regions on chromosomes 5 and 15. Also, possible gene interaction studies are promising avenue for future of genetic COPD research.

It has been postulated that epigenetic modifications are in part driven by genetic variants.[3] In attempt to integrate multiple -omics layers, I focused on the effects of

known COPD genetic variants from chromosomes 15q25.1 (**Chapter 2.2**) and 19q13.2 (**Chapter 2.3**) on genome-wide DNA methylation in blood and gene expression in blood and lungs. Overall, findings of these two chapters highlight the need for integration of multi-omics data to further understand the role of GWAS loci.

In **Chapter 2.2** I performed meQTL and eQTL study and showed evidence suggesting that genetic variations underlying genes in the chromosome 15q25.1 region (in *IREB2, HYKK* and *CHRNA3*) influence COPD susceptibility through changes in blood DNA methylation of *IREB2, PSMA4* and *CHRNA3*. Furthermore, I observed the association of the same variants with several *cis* and *trans* gene expression changes in the lung tissue. These results suggest a disease model in which COPD risk allele of *CHRNA3* variant lowers DNA methylation and increases expression of *IREB2* in COPD patients. We suggest mediation by DNA methylation levels in this region, but future studies using the data from the same tissue should confirm this hypothesis. Even though genetic variants in chromosome 15q25.1 (encompassing the nicotinic receptor genes - *CHRNA3, CHRNA5* and *CHRNB4*) were also previously associated with smoking and smoking has a known effect on DNA methylation, our findings were independent of smoking, making the mediation by smoking not a necessary mechanism driving the relation between DNA methylation and COPD. Both smoking and genetic determinant may lead independently to differential DNA methylation. However, targeted studies in lung tissue should verify these findings.

Similarly, **Chapter 2.3** shows that the top COPD GWAS variant in chromosome 19q13.2 region (rs7937 in *RAB4B, EGLN2*) is associated with lower blood DNA methylation of *EGLN2*, independent of smoking and of COPD. I further showed that this DNA methylation site in *EGLN2* is associated with COPD, again independent of smoking. Having both DNA methylation and expression data in blood, I performed a mediation analysis and showed that differential DNA methylation mediates 42% of the association between the genetic variant and differential expression of *EGLN2*. I also showed the effect of the variant on *cis* and *trans* gene expression changes in lung tissue of *NUMBL, AK097370 (EGLN2), LOC101929709, DNMT3A* and *PAK2*. Our findings are in line with our hypothesis that the life-long change in the DNA methylation is involved in the pathogenesis and onset of COPD in older age, yet further longitudinal studies are needed, testing this hypothesis in large set of lung tissue characterized for multi-omics data.

As smoking plays an important role in the disease development and has a known effect on epigenetic landscape, it is crucial to take this determinant into account in epidemiological studies of COPD.[3,4] In lack of better assessment, self-reported current smoking status (current-, ex- and never-smokers) is widely used variable. To truly remove smoking effects from the equation, the analyses should be performed

in never-smokers only. However, such stratification would significantly reduce the statistical power and the analysis would require large sample sizes to show the real effects. With the aim to investigate DNA methylation signature of airflow obstruction, independent of smoking effects, I performed the largest to date EWAS on lung function levels in never smokers only, presented in **Chapter 2.4**. I identified 36 DNA methylation sites that were highly significantly associated with FEV_1/FVC. This chapter importantly contributes to the current understanding of epigenetic changes in COPD, when smoking effect is excluded. Moreover, this chapter presents current literature on blood DNA methylation in COPD and smoking and shows that the majority of the identified 36 DNA methylation sites are unique for never smokers. DNA methylation of *KLHL32* and *LTV1* genes, among others, indeed may play a role in the disease development in subjects with COPD that never smoked. Although we see associations in blood, many of the methylated genes are expressed in the lung tissue. Genetic variants in *KLHL32* have been associated with post-bronchodilator FEV_1 and FEV_1/FVC in COPD.[5] *LTV1* is shown to be an enhancer of two genes identified in GWAS of FEV_1/FVC response to bronchodilators (*PLAGL1*)[5] and lung cancer (*PHACTR2*)[6] in GeneHancer database.[7] These findings propose a possible new regulatory pathway, independent of smoking, through which DNA methylation of *LTV1* influences genetic susceptibility of COPD. Of course, this is just a hypothesis based on our findings and future functional studies of causality should investigate it further.

Chapter 3 of this thesis is investigating comorbidities of COPD which are known to influence the severity of the disease.[8] Asthma is considered the most common comorbidity, being both a risk factor for COPD and a co-existing disease in the elderly.[9,10] There is a significant overlap between genetic risk factors of asthma and COPD, but the regulatory mechanisms are unknown.[11] It is shown that lower lung function in childhood with subsequent normal or accelerated decline had increased risk of COPD.[12] Early life allergic diseases, lung infections, parental asthma, and maternal smoking predicted worse lung function, while personal smoking amplified the effect of maternal smoking.[12] In **Chapter 3.1** we hypothesized that umbilical cord blood DNA methylation influences childhood lung function, lung development and gene expression, and increases the risk of asthma and COPD in later life. We performed large EWAS meta-analyses of FEV_1, FEV_1/FVC and Forced Expiratory Flow at 75% of FVC ($FEF_{75)}$. EWAS results were pooled into differentially methylated regions (DMRs) and 59 such DMRs in neonatal cord blood were associated with childhood lung function. Multiple DMRs were additionally associated with childhood asthma, adolescent and adult lung function, COPD and with differential gene expression. These findings suggest that epigenetic changes during foetal life might modify the risk of respiratory diseases across the life course. As epigenetic landscape of a foetus is highly

influenced by environment and maternal behaviour,[13] identifying and reducing risk factors in pregnancy, may become a future strategy for prevention of lung diseases.

As COPD is a systemic disease, there are manifestations beyond the airflow obstruction, such as systemic inflammation, muscle degeneration and oxidative stress, which can result in a specific plasma biomarkers profile.[14–16] The identification of these specific biomarker changes can identify or differentiate disease phenotypes even in the early stages of COPD.[17] Therefore, in **Chapter 3.2** I performed a hypothesis-free analysis and identified and replicated the association of Glycoprotein acetyls (GlycA) with COPD, the only association that passed the multiple testing correction. The Rotterdam study and other epidemiological follow-up studies show that GlycA, as measured before the disease onset, increased the risk of COPD and may thus be a risk factor of the disease rather than a consequence. To validate the causal pathway, I further performed Mendelian Randomization (MR) analysis. This analysis clearly showed that the genes driving the risk of COPD, are also associated with GlycA. No evidence was found for the risk factor model in which the genes driving GlycA associate significantly to COPD. This finding suggest that GlycA is an early marker of COPD pathology. It is of interest that **Chapter 3.2** shows that GlycA which is a marker of the acute phase response, also strongly associates to smoking and may thus be a part of an inflammation pathway linking smoking to COPD. However, further functional studies should investigate the specific role of GlycA in COPD pathogenesis, prognosis, severity and treatment response.

Aside from asthma and other pulmonary conditions, many different comorbid diseases add to the burden of COPD. In attempt to identify common pathophysiology and decipher the co-occurrence of those diseases, I performed the genetic correlation analysis of COPD with 126 diseases available on the LD hub database, presented in **Chapter 3.3**. I describe the significant correlation of neuro-psychiatric and cardio-metabolic pathology, as well as female reproductive conditions autoimmune diseases of the bowel and aging disorders such as cataract. Of note is, for the first time shown, significant correlation of coronary artery disease, acute myocardial infarction, angina pectoris, hypertension, diabetes, chronic kidney disease, attention deficit hyperactivity disorder, schizophrenia, family history of depression (depression in sibling) and suggestive correlation with major depressive disorder.

The understudied comorbidity of COPD is depression, investigated in **Chapter 3.4.** As genetic risk factors of depression and COPD do not seem to overlap strongly (**in Chapter 3.3**), I studied different omics layers to try to identify common mechanisms explaining this co-occurrence. Depression related DNA methylation changes have poorly been studied, hence, we have performed largest to date EWAS of depressive symptoms. This chapter presents the identified DNA methylation sites annotated

to the *CDC42BPB* gene, *ARHGEF3* gene, and one intergenic region on chromosome 15q26.1 locus. All three findings point towards axon guidance as the common disrupted pathway in depression. DNA methylation site in *CDC42BPB* gene was also associated with inflammation,[18] smoking status[3] and pack-years of smoking[19] in two independent studies. It has been speculated that *CDC42BPB* may be a future biomarker of COPD[20] and *CDC42BPB* is a downstream target of *CDC42*, whose expression is altered in obese children with astma.[21] At the genetic level, *CDC42BPB* is associated with gamma-glutamyl transferase (http://atlas.ctglab.nl/phewas), which have been found to be differentially expressed in COPD mouse models.[22] Identifying depression related DNA methylation sites was a first step towards unravelling the complex mechanisms underlying depression, but may also shed light on the co-occurrence of depression and COPD and the role of smoking. Future studies should investigate the role of identified genes in COPD and other way around. Using multi-omics approach for investigating comorbidities has a potential to disentangle these relations and provide better treatment options and prognosis for both diseases.

METHODOLOGICAL CONSIDERATIONS

All studies in this thesis were performed using participants' data from Rotterdam Study, a population-based cohort study consisting of 45 years or older people form Rotterdam. In addition, genetic linkage study (**Chapter 2.1**) and metabolic study (**Chapter 3.2**) used data from the Erasmus Rucphen Family study, a genetic isolate from southwest of the Netherlands. Aside from the two studies we also used several Dutch and international studies as part of collaborative efforts of big consortiums. Details and methodological issues of each study are discussed in each chapter of this thesis. Here I would like to mention several general issues important to be considered.

Smoking assessment

Tobacco smoking is one of the major risk factors for many chronic diseases and different types of cancers (**Figure 1**). Therefore, it is one of the most investigated risk factors in epidemiological studies, yet the assessment of smoking has not been standardized. The effects of smoking are commonly assessed using self-administered questionnaires and studied as different variables: smoking behaviour (current smoking status, ever smoking status), smoking quantity (pack-years of smoking, cigarettes per day), nicotine dependence (time from waking up until first cigarette),[23] age of smoking initiation, smoking cessation, second-hand smoke and other. Also, tobacco is used in different ways: in form of cigarettes (different brands

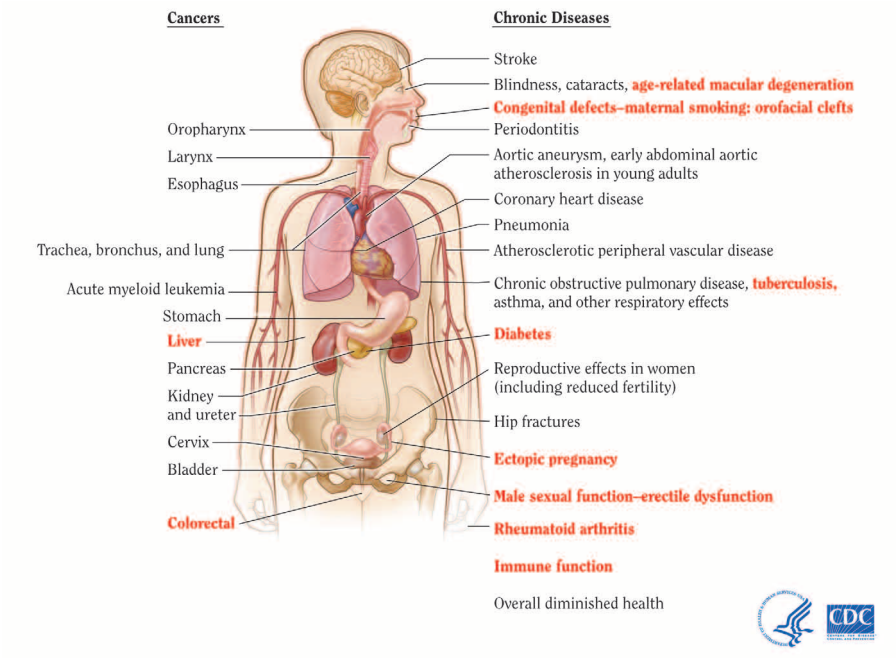

Figure 1. Diseases causally linked to smoking. Source: US Centers for Disease Control and Prevention (https://www.cdc.gov/). In red: a new disease causally linked to smoking in the 2014 Surgeon General's Report: The Health Consequences of Smoking—50 Years of Progress.

vary in the amount of nicotine and noxious particles), smoked in pipes and chewed. Using self-administered questionnaires to assess very heterogeneous exposure may introduce information bias, due to the erroneous classification of subjects. Specifically, patients with lung diseases are shown to likely falsely report their smoking behaviour.[24] The reason for this is often inaccurate memory of past events (recall bias) or deliberately understating as people tend to diminish the adverse behaviour.[25] This leads to frequent underestimation of smoking rates and, when used as a confounder in regression analysis, to over- or underestimation of effect estimates. This is especially important in epigenetic and transcriptomic studies of lung diseases as DNA methylation and gene expression is known to be affected by smoking.[3,26] Genetic studies have been identified risk loci associated with smoking, such as nicotinic receptors on chromosome 15, studied in **Chapter 2.2** of this thesis. In attempt to exclude the smoking effect, in **Chapter 2.4** I study DNA methylation in never smokers only. In **Chapter 3.2** I investigate the role of smoking in the identified COPD-GlycA associations. When studying multi-omics of lung diseases, where smoking plays an important role, it is imperative to take real smoking ef-

fects into account. However, in most of the cohorts used in this thesis, in absence of a better assessment, smoking behaviour is measured using self-administered questionnaires as current smoking status (current, ex or never smokers) and pack-years of any smoking (including cigarettes and pipes). Although self-administered questionnaire is quick, easy and cheap way of assessing smoking and thus widely used, the question remains whether it captures the real smoking exposure. Objective assessment of smoking biomarkers, such as nicotine, cotinine and the exhaled carbon monoxide concentrations, is available but is more expensive and less used in population-based studies.[27] These methods would undoubtedly report actual current smoking exposure, even in never-smokers exposed to second-hand smoke, but would not show difference between never-smokers and ex-smokers.[28] It is shown that there is a big discrepancy between self-reported and objectively measured current smoking status.[24,29] While the next generation technologies for assessment of omics data are developing fast and large datasets are readily available through national biobanks and large consortia, phenotyping of smoking exposure is still plummeted with misclassification of subjects. Therefore, it is of utmost importance for objective smoking assessment methods to become more widely used, especially in the next-generation multi-omics studies of lung diseases.

DNA methylation assessment

All cohorts included in the epigenetic studies of this thesis were quantifying DNA methylation using Infinium Human Methylation 450 Bead Chip from llumina, the array with more than 450 thousand CpG sites. It is known that 450K array measures only 1.7% of all CpGs mostly covering CpG-islands.[30] CpG island is a stretch of DNA with the highest frequency of CpGs, mostly located in the gene promotor regions and gene bodies. Therefore, when analysing data, limited genome coverage should be considered, as enhancer regions are almost completely missing. Furthermore, it has been discovered that some hybridization probes of 450K array co-hybridize to similar genomic sequences (cross-reactive CpGs) or target CpGs that overlap with genetic polymorphisms (polymorphic CpGs), so the measured methylation levels may reflect the underlying SNPs.[31] Potentially cross-reactive and polymorphic CpGs can cause a measurement bias and have been annotated previously.[31] This can confound the data, though unknown to which extent, and found associations should be interpreted with caution. Being a technical issue, adjusting for the underlying SNPs would not correct this error, while excluding the biased sites would possibly generate false-negative findings. To be stringent and sure of the validity of our findings, in most of our epigenetic studies we have excluded both cross-reactive and polymorphic CpGs. Future studies using the new Illumina 850K EPIC array, in which

problematic CpGs are corrected and many more added, covering the enhancer regions, this issue should be solved.[32]

Tissue of interest

For the quantification of the multi-omics data, blood is the most commonly used tissue as it is easily obtained and cheap. However, when using blood in studying epigenetics, transcriptomics and metabolomics of lung diseases the question arises whether changes identified in blood represent the processes in the tissue of interest – the lungs. On the other hand, COPD represents more of a syndrome than a single disease including different phenotypes and underlying processes which we are still trying to comprehend. Those are airway obstruction in bronchitis, loss of lung parenchyma in emphysema, as well as systemic effects such as the inflammation and oxidative stress as a response to the noxious particles, muscle wasting and changes in metabolism.[33] It has been shown that those systemic effects are detectable in plasma e.g. through the role of the macrophage and neutrophils in the pathogenesis of COPD (**Figure 2**),[34] justifying the use of blood as the tissue of interest.[35,36] In this

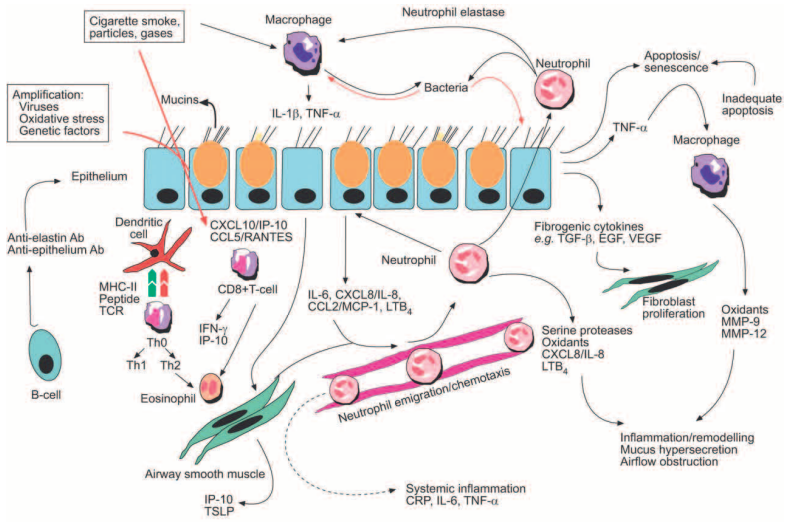

Figure 2. Summary of inflammatory and cellular interactions linking chronic cigarette exposure to the chronic inflammation of chronic obstructive pulmonary disease (COPD). Ab: antibody; Th: T-helper cell; MHC: major histocompatibility complex; TCR: T-cell receptor; CXCL: CXC chemokine ligand; IP: interferon (IFN)-γ-inducible protein; CCL: CC chemokine ligand; RANTES: regulated on activation, normal T-cell expressed and secreted; TSLP: thymic stromal lymphopoietin; IL: interleukin; TNF: tumour necrosis factor; MCP: monocyte chemotactic protein; LT: leukotriene; CRP: C-reactive protein; TGF: transforming growth factor; EGF: epidermal growth factor; VEGF: vascular endothelial growth factor; MMP: matrix metalloproteinase. Source: Chung KF, Adcock IM. *Eur Respir J*. 2008;31(6):1334-1356. doi:10.1183/09031936.00018908.

thesis blood was used for quantification of DNA methylation levels as well as gene expression and metabolic levels. However, for gene expression analyses described in **Chapters 2.2** and **2.3** we confirmed the regulation of gene expression by the COPD-associated genetic variants in lung tissue. When integrating multi-omics levels, it is very important to have all the data in lung tissue so that causal relationships involved in COPD pathophysiology can be investigated.

An important drawback to be discussed about the use of expression in blood is that it is quite heterogeneous tissue, consisting of many cell types, whose proportions may vary in different people and with disease severity. In whole blood DNA methylation studies, such as those described in this thesis, different types of leukocytes are used (lymphocytes, monocytes and granulocytes). As DNA methylation can be cell-specific, using a mix of cell types prone to inter-individual variability can give biased results.[37] Therefore, these analyses should be adjusted for cell proportions, either measured within the study or predicted based on the DNA methylation data using computational methods.[38] In the chapters of this thesis we have used both measured cell proportions and estimated proportions utilizing Houseman method.[39] However, it is important to consider that, while Housman estimation is good method to use for consistency in the big consortium efforts, it is based on half a dozen reference individuals and may not completely represent the proportions in the general population.

Causality

In this thesis we often performed cross-sectional association analyses, disabling us to claim any causal relationships between the studied exposure and the outcome. Even the variants identified in linkage analysis of **Chapter 2.1** may not be causal for COPD but simply in a linkage disequilibrium with the causal variant which is not genotyped. The fact that the identified genes are expressed in the lung tissue, may suggest that the real causal variant is within the gene. However, any claims of causality should be taken with caution and confirmed by functional studies using knockout models. On the other hand, DNA methylation is a dynamic process subjected to change and influenced by both external and internal effects. In this thesis I have reported associations of GWAS SNPs with blood DNA methylation, further associated with COPD (**Chapters 2.2** and **2.3**), yet it may be that having COPD changes DNA methylation at those sites. I have also showed association of new-born cord blood DNA methylation with lung function and asthma and COPD in later life in **Chapter 3.1.** In this chapter, using different time points for measuring the data could suggest that reverse causality may not be possible, but this should be confirmed in future longitudinal studies on the same people. In **Chapter 2.3,**

having both DNA methylation and gene expression data derived from blood, I could perform mediation analysis and I showed that DNA methylation of *EGLN2* mediates 42% of the association between rs7937 and expression of *EGLN2.* This may suggest a direction of the effect, but the reverse causality may also happen, as differential gene expression can affect DNA methylation levels of *EGLN2*.

The Holy Grail for testing causality in observational epidemiological research is to conduct (nested) follow-up studies in which the exposure is measured before the disease. We conducted a nested follow-up study of various metabolites and COPD (**Chapter 3.2**) and found evidence that that Glycoprotein acetyls (GlycA) were associated with an increased risk of COPD. Recent development in genetics have made it possible to use genes as instrumental variable (IV) to test for causality using Mendelian Randomisation (MR; explained in **Chapter 1**). In **Chapter 3.2** I tested the most likely causal pathway underlying the association of COPD with GlycA. I showed in this chapter that COPD is causally related to the elevated levels of GlycA, rather than other way around. Although the conclusion is important, it is important to consider the possibility that the assumptions underlying MR are violated. These include:

1. The IV has to be associated with the exposure;
2. The IV has to be independent of any confounders of the exposure-outcome association and
3. The IV has to be related to the outcome only through the exposure.

However, as genetic variants used as IV may associate with other unknown confounders of the COPD-GlycA association, the condition two may have been violated. The finding that the genes predicting GlycA do not predict COPD makes it unlikely that GlycA is a risk factor. However, the hypothesis that GlycA is a biomarker of COPD pathology remains to be studied further prospectively.

External validity

External validity, the extent to which the results can be generalized to other situations and to other people, is an important factor to consider in epidemiological study. This issue should be taken into account when interpreting the results of **Chapter 2.1** of this thesis. Identified rare variants, identified in the genetic isolate, may be population-specific thus rare or non-existent in other populations.[40] This especially stands for the two variants which we could not replicate in the population-based setting. Furthermore, the replication cohorts were mainly of European ancestry, questioning the generalizability to other ancestries, since it is known that genetic variants have different effects in different ethnic populations. Utilizing large datas-

ets to replicate our findings is required before speculating on the external validity of these findings.

Differentially methylated regions

It has been postulated that differential DNA methylation at a single CpG site has a small effect on the risk for the disease and it should be evaluated considering the effects of the neighbouring sites or the whole region,[41] which makes sense from the biological point of view. These differentially methylated regions (DMR) are estimated using Comb-p method in **Chapter 3.1** of this thesis.[42] This method calculates auto-correlation, combines neighbouring P-values, corrects for false discovery rate, finds regions of enrichment and assigns a P-value to those regions.[42] The number of identified DMRs is much larger than identified single sites since this method increases power to identify region of interest. We identified 59 DMRs associated with childhood lung function, of which 31% were associated with childhood asthma, 19% with adolescent lung function, 15% with adult lung function, and 15% with COPD while 54% were influencing gene expression in childhood and 31% in adulthood. However, as Illumina 450K array is known to have unequable coverage of CpG sites throughout the genome, this method has its limitations. Therefore, simply pooling the CpG sites using computational methods, without any a-priori knowledge is questionable and the results should be interpreted with caution. The use of EPIC 850K array in the future, which has much better coverage of CpGs across the genome, may bridge this limitation.

POTENTIAL IMPLICATIONS AND FUTURE DIRECTIONS

Clinical implications and directions for future research were discusses in detail in every chapter of this thesis. Here I will highlight my most important ideas for the future.

Development of technical means which can withstand computationally demanding analyses opened new avenues for research of complex diseases. COPD, as a complex disease, already benefitted from the use of hypothesis-generating -omics approaches. GWAS identified over 40 loci associated with either lung function or COPD, yet the reproducibility of these findings is very low and the heritability is still largely missing.[43] GWAS variants are usually common with very small effect size. Using stringent, genome-wide significant thresholds we will need to increase sample sizes to reach enough statistical power to discover new common genetic variants. In the future, this can be done using large biobanks, collections of nation-wide data,

lately developing in several countries. It will be crucial to include information on different ancestries, as currently most of the findings are identified on Caucasians, and is has been speculated that this is leading to a disparity in future health care. Worldwide collaborations would also facilitate these efforts through large consortiums such as CHARGE and PACE consortiums, used in **Chapter 3** of this thesis. In order to confirm the role of rare variants on chromosome 11, described in **Chapter 2.1,** and to generalize findings to the general population, larger candidate-gene studies will be useful. Furthermore, large studies utilizing whole-exome and -genome sequencing should investigate the role of linked regions in chromosomes 5 and 15 in COPD. Such studies are now feasible and affordable at a large scale for instance the UK biobank. Gene-gene and gene-environment interactions should be investigated in further attempt to explain the missing heritability in the future. There has been little success up until now and this field awaits advances through upscaling the size of studies and the development of new computational approaches (e.g. deep learning). Functional studies are further needed to definitely confirm the role of novel variants in the disease pathogenesis.

Further, we show usefulness of investigating metabolic profiles of the disease to hopefully differentiate molecular processes in specific tissues. I found that GlycA is a biomarker of early COPD pathology that is elevated before the diagnosis of COPD and future studies widening the net of metabolites studied are likely to find other metabolites, improving the prediction of disease and yielding new information that in combination can empower future (preventive) trials.

In the future, multi-omics studies of COPD would benefit from longitudinal design, measuring multiple omics layers in the lung tissue of the same people in multiple time points. Basing the study in well characterized epidemiological cohorts will allow to remove the confounding effects of smoking and medication. This would lead to better inferences on causality and direction of the effects. MR is a useful method to infer causality and disentangle complex relations of multiple omics layers and should be used in the future research. Such studies would also allow to derive objective smoking assessment based on methylation, transcription and metabolomics profiles. Similarly, one can aim to capture the effect of air pollution and other risk factor of COPD in omics signatures.

Improved and more discrete phenotyping of COPD and its confounders, such as smoking behaviour, is required in order to understand the genetic architecture of the disease. This includes investigating comorbidities of COPD and their shared and overlapping risk factors. As comorbidities are influencing disease severity, prognosis and treatment response and complete well-being of the patient, it is imperative to determine to what extend these can be prevented.

Large scale genome-wide multi-omics studies in the lung tissue are the next phase in the respiratory research, integrating data in the context of a network medicine.[44] This may improve understanding of the disease heterogeneity, improve classification and identification of individuals in high risk and translate the findings to clinical care and prevention (**Figure 3**). This may open new avenues for precision medicine in the future.[45]

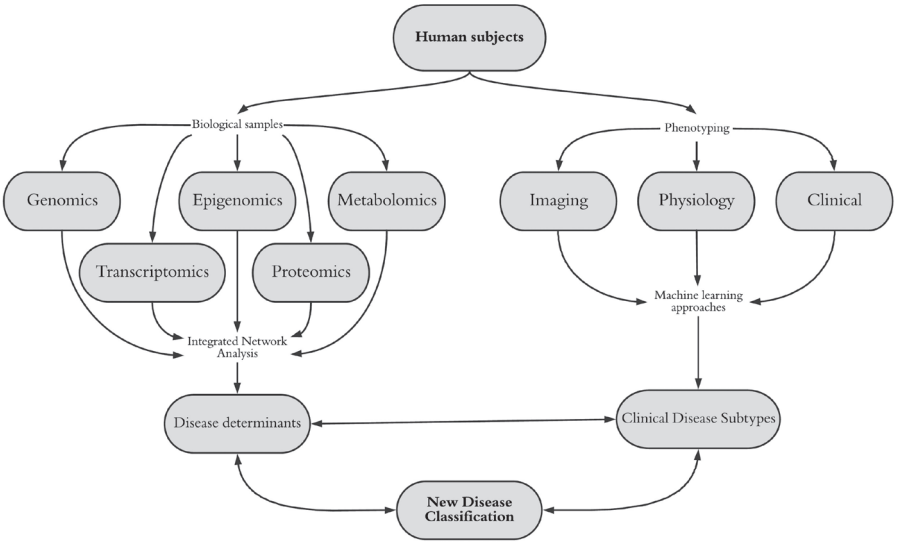

Figure 3. Potential Approaches to Reclassify Complex Diseases in Network Medicine. Adapted from Silverman and Loscalzo *Discov Med*. 2012;14(75):143-152.

The final aim of the network medicine, the identification of important disease determinants and reclassification of complex disease, such as COPD, is to have novel drug development strategies in the future and improve clinical care (**Figure 4**).[46,47]

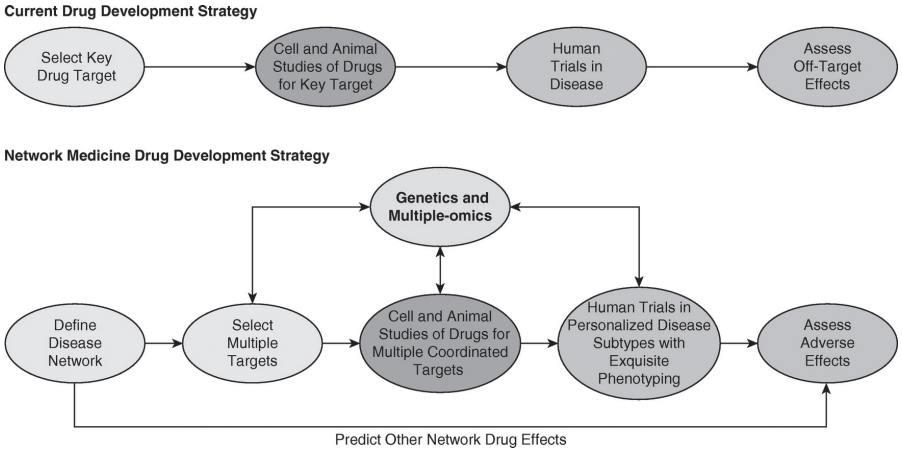

Figure 4. Current and Network Medicine Approaches to Drug Development for Complex Diseases. Adapted from Silverman EK, Loscalzo J. Clin Pharmacol Ther. 2012;93(1):26-8.

Conclusion

In this thesis I studied COPD using several omics layers in attempt to elucidate the molecular mechanisms underlying the disease. I have identified several novel risk factors for COPD and its comorbidities, proposed regulatory pathways and highlighted the need for integration of the multi-omics data. Furthermore, I have discussed methodological challenges and ideas from which the future respiratory research could benefit. Findings of this thesis require functional confirmation, but I am confident that it represents another important step on a path towards improved clinical care and prevention of COPD, based on precision medicine.

REFERENCES

1. Lozano R, Naghavi M FK. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380(380):2095–2128. doi:https://doi.org/10.1016/S0140-6736(12)61728-0
2. European Respiratory Society. European lung white book. http://www.erswhitebook.org/chapters/chronic-obstructive-pulmonary-disease/. Published 2016.
3. Yet I, Tsai P-C, Castillo-Fernandez JE, Carnero-Montoro E, Bell JT. Genetic and environmental impacts on DNA methylation levels in twins. *Epigenomics*. 2016;8(1):105-117. doi:10.2217/epi.15.90
4. Joehanes R, Just AC, Marioni RE, et al. Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet*. 2016;9(5):436-447. doi:10.1161/CIRCGENETICS.116.001506
5. Lokke A, Lange P, Scharling H, Fabricius P, Vestbo J. Developing COPD: a 25 year follow up study of the general population. *Thorax*. 2006;61(11):935-939. doi:10.1136/thx.2006.062802
6. Lutz SM, Cho MH, Young K, et al. A genome-wide association study identifies risk loci for spirometric measures among smokers of European and African ancestry. *BMC Genet*. 2015;16(1):138. doi:10.1186/s12863-015-0299-4
7. Wang L-E, Gorlova OY, Ying J, et al. Genome-Wide Association Study Reveals Novel Genetic Determinants of DNA Repair Capacity in Lung Cancer. *Cancer Res*. 2013;73(1):256-264. doi:10.1158/0008-5472.CAN-12-1915
8. Fishilevich S, Nudel R, Rappaport N, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database*. 2017;2017. doi:10.1093/database/bax028
9. Chatila WM, Thomashow BM, Minai OA, Criner GJ, Make BJ. Comorbidities in Chronic Obstructive Pulmonary Disease. *Proc Am Thorac Soc*. 2008;5(4):549-555. doi:10.1513/pats.200709-148ET
10. Silva GE. Asthma as a Risk Factor for COPD in a Longitudinal Study. *Chest*. 2004;126(1):59-65. doi:10.1378/chest.126.1.59
11. Aryal S, Diaz-guzman E, Mannino DM. Asthma Treatment Options in Asthma and Chronic Obstructive Pulmonary Diseases Overlap Syndrome. *Touch Briefings, Eur Respir Dis*. 2011;7(2):101-105.
12. Hobbs BD, De Jong K, Lamontagne M, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet*. 2017;49(3):426-432. doi:10.1038/ng.3752
13. Bui DS, Lodge CJ, Burgess JA, et al. Childhood predictors of lung function trajectories and future COPD risk: a prospective cohort study from the first to the sixth decade of life. *Lancet Respir Med*. 2018;6(7):535-544. doi:10.1016/S2213-2600(18)30100-0
14. Eckhardt F, Lewin J, Cortese R, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet*. 2006;38(12):1378-1385. doi:10.1038/ng1909
15. Chen Q, Deeb RS, Ma Y, Staudt MR, Crystal RG, Gross SS. Serum metabolite biomarkers discriminate healthy smokers from COPD smokers. *PLoS One*. 2015;10(12):e0143937. doi:10.1371/journal.pone.0143937
16. Adamko DJ, Nair P, Mayers I, Tsuyuki RT, Regush S, Rowe BH. Metabolomic profiling of asthma and chronic obstructive pulmonary disease: A pilot study differentiating diseases. *J Allergy Clin Immunol*. 2015;136(3):571-580.e3. doi:10.1016/j.jaci.2015.05.022
17. Ubhi BK, Riley JH, Shaw PA, et al. Metabolic profiling detects biomarkers of protein degradation in COPD patients. *Eur Respir J*. 2012;40(2):345-355. doi:10.1183/09031936.00112411

18. Nobakht M Gh BF, Aliannejad R, Rezaei-Tavirani M, Taheri S, Oskouie AA. The metabolomics of airway diseases, including COPD, asthma and cystic fibrosis. *Biomarkers Biochem Indic Expo response, susceptibility to Chem*. 2015;20(1):5-16. doi:10.3109/1354750X.2014.983167
19. Ligthart S, Marzi C, Aslibekyan S, et al. DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. *Genome Biol*. 2016;17(1):255. doi:10.1186/s13059-016-1119-5
20. de Vries M, van der Plaat DA, Nedeljkovic I, et al. From blood to lung tissue: effect of cigarette smoke on DNA methylation and lung function. *Respir Res*. 2018;19(1):212. doi:10.1186/s12931-018-0904-y
21. Leidinger P, Keller A, Heisel S, et al. Novel autoantigens immunogenic in COPD patients. *Respir Res*. 2009;10(1):20. doi:10.1186/1465-9921-10-20
22. Rastogi D, Nico J, Johnston AD, et al. CDC42-related genes are upregulated in helper T cells from obese asthmatic children. *J Allergy Clin Immunol*. 2018;141(2):539-548.e7. doi:10.1016/j.jaci.2017.04.016
23. Zhang H, Sun D, Li D, et al. Long non-coding RNA expression patterns in lung tissues of chronic cigarette smoke induced COPD mouse model. *Sci Rep*. 2018;8(1):7609. doi:10.1038/s41598-018-25702-3
24. Transdisciplinary Tobacco Use Research Center (TTURC) Tobacco Dependence TB, Baker TB, Piper ME, et al. Time to first cigarette in the morning as an index of ability to quit smoking: implications for nicotine dependence. *Nicotine Tob Res*. 2007;9 Suppl 4(Suppl 4):S555-70. doi:10.1080/14622200701673480
25. Stelmach R, Fernandes FLA, Carvalho-Pinto RM, et al. Comparison between objective measures of smoking and self-reported smoking status in patients with asthma or COPD: are our patients telling us the truth? *J Bras Pneumol*. 2015;41(2):124-132. doi:10.1590/S1806-37132015000004526
26. Sponsiello-Wang Z, de La Bourdonnaye G, David M, Lüdicke F, Weitkunat R. Accuracy of the Smoking Questionnaire. *Beiträge zur Tab Int to Tob Res*. 2017;27(8):224-239. doi:10.1515/cttr-2017-0023
27. Hackett NR, Butler MW, Shaykhiev R, et al. RNA-Seq quantification of the human small airway epithelium transcriptome. *BMC Genomics*. 2012;13(1):82. doi:10.1186/1471-2164-13-82
28. SRNT Subcommittee on Biochemical Verification NL, Jacob P, Ahijevych K, et al. Biochemical verification of tobacco use and cessation. *Nicotine Tob Res*. 2002;4(2):149-159. doi:10.1080/14622200210123581
29. Baltar VT, Xun WW, Chuang S-C, et al. Smoking, secondhand smoke, and cotinine levels in a subset of EPIC cohort. *Cancer Epidemiol Biomarkers Prev*. 2011;20(5):869-875. doi:10.1158/1055-9965.EPI-10-1235
30. Hald J, Overgaard J, Grau C. Evaluation of Objective Measures of Smoking Status A Prospective Clinical Study in a Group of Head and Neck Cancer Patients Treated with Radiotherapy. *Acta Oncol (Madr)*. 2003;42(2):154-159. doi:10.1080/02841860310005020
31. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium Methylation 450K technology. *Epigenomics*. 2011;3(6):771-784. doi:10.2217/epi.11.105
32. Chen YA, Lemire M, Choufani S, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*. 2013;8(2):203-209. doi:10.4161/epi.23470

33. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics*. 2016;8(3):389-399. doi:10.2217/epi.15.114
34. Wouters EFM, Wouters. Chronic obstructive pulmonary disease * 5: Systemic effects of COPD. *Thorax*. 2002;57(12):1067-1070. doi:10.1136/thorax.57.12.1067
35. Chung KF, Adcock IM. Multifaceted mechanisms in COPD: inflammation, immunity, and tissue repair and destruction. *Eur Respir J*. 2008;31(6):1334-1356. doi:10.1183/09031936.00018908
36. Rahman I, Morrison D, Donaldson K, MacNee W. Systemic oxidative stress in asthma, COPD, and smokers. *Am J Respir Crit Care Med*. 1996;154(4):1055-1060. doi:10.1164/ajrccm.154.4.8887607
37. NOGUERA A, BUSQUETS X, SAULEDA J, VILLAVERDE JM, MacNEE W, AGUSTÍ AGN. Expression of Adhesion Molecules and G Proteins in Circulating Neutrophils in Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med*. 1998;158(5):1664-1668. doi:10.1164/ajrccm.158.5.9712092
38. Houseman EA, Kim S, Kelsey KT, Wiencke JK. DNA Methylation in Whole Blood: Uses and Challenges. *Curr Environ Heal Reports*. 2015;2(2):145-154. doi:10.1007/s40572-015-0050-3
39. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol*. 2014;15(2):R31. doi:10.1186/gb-2014-15-2-r31
40. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13(1):86. doi:10.1186/1471-2105-13-86
41. Tennessen JA, Bigham AW, O'Connor TD, et al. Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. *Science (80-)*. 2012;337(6090):64-69. doi:10.1126/science.1219240
42. Jaffe AE, Murakami P, Lee H, et al. Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. *Int J Epidemiol*. 2012;41(1):200-209. doi:10.1093/ije/dyr238
43. Pedersen BS, Schwartz DA, Yang I V., Kechris KJ. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. *Bioinformatics*. 2012;28(22):2986-2988. doi:10.1093/bioinformatics/bts545
44. Ortega VE. Picking the Right Fruit: Intersecting Chronic Obstructive Pulmonary Disease Genome-Wide Association Study Discoveries with Epigenetics. *Am J Respir Crit Care Med*. 2018;197(10):1237-1239. doi:10.1164/rccm.201801-0084ED
45. Silverman EK, Loscalzo J. Network medicine approaches to the genetics of complex diseases. *Discov Med*. 2012;14(75):143-152. http://www.ncbi.nlm.nih.gov/pubmed/22935211. Accessed November 29, 2018.
46. Gligorijević V, Malod-Dognin N, Pržulj N. Integrative methods for analyzing big data in precision medicine. *Proteomics*. 2016;16(5):741-758. doi:10.1002/pmic.201500396
47. Agustí A, Antó JM, Auffray C, et al. Personalized Respiratory Medicine: Exploring the Horizon, Addressing the Issues. Summary of a BRN-AJRCCM Workshop Held in Barcelona on June 12, 2014. *Am J Respir Crit Care Med*. 2015;191(4):391-401. doi:10.1164/rccm.201410-1935PP
48. Silverman EK, Loscalzo J. Developing new drug treatments in the era of network medicine. *Clin Pharmacol Ther*. 2013;93(1):26-28. doi:10.1038/clpt.2012.207