Genes, DNA methylation and exposures underlying COPD in never-smokers
van der Plaat, Dini

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:
2017

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
vан дер Плат, Д. А. (2017). Genes, DNA methylation and exposures underlying COPD in never-smokers [Гронинген]: Рицксиверништ Гронинген

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).

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.
Summary, general discussion and future perspectives
The aim of this PhD project was to assess the mechanisms that play a role in COPD in never-smokers, including genes, DNA methylation and environmental exposures. Chapters 2 to 4 have shown that several genetic loci, i.e. *HHIP*, *FAM13A*, *MECR*, *FABP7* and *NFYC*, are susceptibility genes for the development of large or small airway obstruction in specifically never-smokers and thus have an effect regardless of exposure to cigarette smoking. In chapter 5 we have shown that associations between DNA methylation levels and COPD were not identified in never- or current-smokers and in chapter 6 we have shown that air pollution is associated with restrictive ventilatory patterns in the general population. Chapters 7 to 9 have shown that altered DNA methylation levels related to air pollution and occupational exposure to gases/fumes, mineral dust, biological dust, or pesticides could be the mechanism how these exposures affect lung function in the general population or in never-smokers specifically.

**Genetics of COPD in never-smokers**

**Chapter 2** describes a genome-wide association study in which we assessed associations between SNPs and lung function levels of FEV$_1$ and FEV$_1$/FVC in 5,070 never-smokers of LifeLines. We were able to replicate associations between FEV$_1$/FVC and two SNPs annotated to the well-known COPD genes *HHIP* and *FAM13A* in a meta-analysis of four independent cohorts (Vlagtwedde-Vlaardingen (n = 432) and Rotterdam study I-III (n = 408, 379, 747)). The two SNPs were also associated with gene expression levels of *HHIP* and *FAM13A* in lung tissue. In addition, the genetic risk score analysis showed that subjects carrying all four risk alleles of the two independent SNPs have a 2.4% lower FEV$_1$/FVC level, a credible clinically relevant effect. Of interest, a substantial proportion of our study population, i.e. 6%, was carrying all four risk alleles. *HHIP* is known to play an important role in fetal lung branching development and silencing of *HHIP* in a bronchial epithelial cell line was shown to lead to differential expression of about 300 genes enriched for cell growth, lung extracellular matrix and genes associated with COPD. *FAM13A* may play a role in emphysema development, since *FAM13A* null mice were less susceptible to develop emphysema suggestively due to increased cell proliferation and activation of b-catenin signaling. We concluded, based on our findings in never-smokers and data from the literature, that the genes *HHIP* and *FAM13A* confer a risk for airway obstruction in general and that it is not driven exclusively by cigarette
smoking. Moreover, these two genes might have an impact on the prevalence of COPD worldwide, especially in countries where COPD in never-smokers is more prevalent.

In chapter 3 a genome-wide association study is described in which we assessed associations between SNPs and small airways function in the same 5,070 never-smokers of LifeLines, using the spirometric measurement of FEF$_{25-75}$. The risk allele of the same SNP annotated to HHIP, the gene known to be involved in lung development, was genome-wide significantly associated with a 98 mL/s lower level of FEF$_{25-75}$ ($p = 7.27 \times 10^{-8}$). In the sensitivity analysis excluding subjects with large airways obstruction (FEV$_1$/FVC $< 70\%$ and FEV$_1$ $< 80\%$pred, n = 4,855), the HHIP SNP was no longer genome-wide significant ($b = 72, p = 2.67 \times 10^{-5}$), but a SNP annotated to MECR was significantly associated with lung function of the small airways ($p = 7.61 \times 10^{-8}$). This MECR SNP was associated with gene expression levels of MECR and EPB41 in lung tissue. MECR is involved in mitochondrial fatty acid synthesis and over-expression of MECR increases PPARα activity, which might play a role in emphysema progression. This study therefore provided novel insights into the role of the well-known COPD gene HHIP in both the small and large airways and showed supportive evidence for the role of the novel gene MECR in specifically small airways obstruction.

Chapter 4 describes genome-wide association studies on airway obstruction, defined as either FEV$_1$/FVC $< 70\%$ or FEV$_1$/FVC $< \text{lower limit of normal (LLN)}$. There is a considerable controversy about which definition should be used in research and clinical practice, since both may lead to misclassifications. Using the $< 70\%$ definition may lead to overdiagnosis in elderly and underdiagnosis in young adults, yet using $< \text{LLN}$ may lead to overdiagnosis in young adults and underdiagnosis in the elderly. So far, GWA studies have shown only a few genetic regions that overlapped, likely due to the use of different airway obstruction definitions (doctor diagnosed, FEV$_1$/FVC $< 70\%$, or $< \text{LLN}$) and different populations under study. We therefore aimed to assess the genetic overlap between the two airway obstruction definitions in the same individuals of our LifeLines cohort, stratified by smoking. We expected a reasonable overlap in associated SNPs between the two definitions, since 96% of the never-smokers and 93% of the ever-smokers were classified in the same way. Surprisingly, only 3% and 6% of the SNPs were overlapping between the two definitions at $p < 10^{-4}$, respectively. At $p < 0.05$, the overlap was still limited, i.e. 21% and 22% respectively. This implies that the definition of airway obstruction and the analytical strategy (discovery-replication design) have a
substantial influence on the GWAS results, and thus on the variants that will be selected for replication. In addition, none of the selected SNPs overlapped of both definitions between never- and ever-smokers at $p < 10^{-4}$, and there was a very weak correlation between the $p$-values and between the ORs. This study therefore also highlighted the importance of stratifying the analysis according to smoking status, since there might be different underlying (genetic) causes between smokers and non-smokers. Finally, our results also suggested that the genes FABP7 and NFYC(-AS1) could play a role in the pathogenesis of airway obstruction in never-smokers.

**DNA methylation and COPD**

In chapter 5 a genome-wide DNA methylation study is described in which we assessed whether DNA methylation levels are associated with COPD (airway obstruction defined as $\text{FEV}_1/\text{FVC} < 70\%$) in a general population stratified by never- and current-smoking. In both the never- and current-smoking analyses, none of the CpGs were genome-wide significantly associated with COPD. This is in contrast with the only study thus far exploring the association between DNA methylation and COPD by Qiu et al. In the latter study, 349 significant CpGs were identified, but we were not able to replicate these results. This difference may be explained by the more severe COPD phenotype of the included patients or the lack of adjustment for white blood cell composition in the study of Qiu et al. In depth analyses showed that the use of different Illumina platforms (27k versus 450K) could not explain the differences between both studies. To conclude, we did not observe any significant associations between DNA methylation levels and the presence of COPD in never- or current-smokers.

**Effects of air pollution on lung function and DNA methylation**

Chapter 6 describes a study assessing the association of air pollutants ($\text{NO}_2$, $\text{PM}_{2.5}$, $\text{PM}_{10}$ and $\text{PM}_{2.5}$ absorbance) with lung function levels ($\text{FEV}_1$, $\text{FVC}$ and $\text{FEV}_1/\text{FVC}$). Annual air pollution exposure of 51,855 LifeLines subjects was estimated at their home address. For all air pollutants the negative (non-significant) estimated effects on the level of $\text{FEV}_1$ were smaller than the negative effects on the level of $\text{FVC}$, which resulted in positive associations with their ratio ($\text{FEV}_1/\text{FVC}$). Therefore this study confirmed results from previous European studies that long-term exposure to low levels of ambient air pollution are associated with restrictive ventilatory patterns rather than obstructive ventilatory patterns. Restrictive ventilatory patterns may reflect pulmonary fibrosis, a disease
which has thus far largely been understudied in relation to air pollution exposure in large-scale epidemiological studies.

**Chapter 7** describes the largest genome-wide DNA methylation study so far assessing associations between long-term ambient air pollution exposure (NO$_2$, PM$_{2.5}$, PM$_{10}$ and PM$_{2.5}$ absorbance) and DNA methylation in 1,017 LifeLines subjects. In addition, we investigated whether methylation mediates the association between air pollution and lung function. We found genome-wide significant associations between NO$_2$ exposure and DNA methylation for seven CpG sites. No significant associations were found for PM exposure. Two CpG sites, annotated to ARF5 and GNG2, were borderline significant mediators in the association between NO$_2$ and FEV$_1$/FVC. Replication of these findings in other cohorts is, however, necessary to elucidate the suggested role of epigenetic variability in the pathogenesis of NO$_2$-exposure related respiratory disease.

**Occupational exposures and DNA methylation**

In **Chapter 8** a genome-wide DNA methylation study is described in which we assessed the association between occupational exposure to gases/fumes, mineral or biological dust and differentially methylated regions (DMRs) stratified by smoking status. In addition, we assessed if these DMRs are associated with gene expression and mediate the association between occupational exposure and lung function. 903 never-smokers and 658 current-smokers of LifeLines were included with either no, low, or high exposure to the three occupational exposures (estimated based on current or last held job). A total of 41 DMRs were significantly associated with occupational exposures in both never- and current-smokers, 16 with gases/fumes, 7 with mineral dust and 18 with biological dust. Using data of the Rotterdam study III-1 (233 never-smokers and 489 ever-smokers), we replicated two DMRs in never-smokers being associated with gases and fumes (VTRNA2–1 and GNAS), and one being associated with mineral dust (CCDC144NL). Several of our identified DMRs showed evidence for mediation of the association between the occupational exposure and lung function, and in addition were associated with gene expression levels in blood (BIOS data, $n = 2,802$). Some regions were associated with two types of occupational exposure and multiple DMRs were located within the transcript start site of gene expression regulating genes. To conclude, we showed that occupational exposures associated with differential methylation of gene expression regulating genes. Our data also suggests that occupational exposures affect lung function via this methylation.
Chapter 9 describes a genome-wide DNA methylation study assessing the association between occupational exposure to pesticides and 420,938 DNA methylation sites. High pesticide exposure was genome-wide significantly (FDR $p < 0.05$) associated with differential methylation of, in total, 26 CpGs that annotated to 23 genes. No CpGs were associated with pesticide exposure in the complete cohort, but 20 of the identified CpGs were found in subjects with airway obstruction. Several of the 23 annotated genes, e.g. \textit{RYR1}, \textit{ALLC}, \textit{PTPRN2}, \textit{LRRC3B}, \textit{PAX2}, and \textit{VTRNA2–1} are genes previously linked to either pesticide exposure or lung related diseases. Six of the identified CpGs were associated with gene expression levels, and thus might have a biological function. We showed for the first time that occupational exposure to pesticides is associated with differential DNA methylation assessed in a hypothesis free genome-wide approach. Replication of these results should indicate whether these CpGs are true findings and whether they play a role in airway disease pathogenesis.
DISCUSSION AND FUTURE PERSPECTIVES

Chronic obstructive pulmonary disease (COPD) is a complex disease with multiple phenotypes, underlying causes and biological pathways. Many studies have been performed to assess the mechanisms that play a role in COPD, but they were performed primarily in smokers or were population-based studies with no special focus on never-smokers. This is not surprising since cigarette smoking is considered to be the most important preventable risk factor for COPD in the developed world. However, the number of COPD patients without a history of smoking is substantial (25–45%) and multiple epidemiological studies have identified many other risk factors associated with COPD, including air pollution and occupational exposures. In addition, never-smoking COPD patients in the developed world seem to have a milder disease (less symptoms, inflammation and comorbidities) and thus might have different underlying pathways compared to smokers with COPD. Interestingly, we see a gradual decline in the number of smokers and an increase of COPD patients in the Netherlands. The proportion of never-smoking COPD patients might thus increase in the future. We therefore tried, in contrast to previous research focusing on smokers, to unravel the underlying mechanisms of COPD in never-smokers and compared the findings to those in smokers. The research described in this thesis focused in particular on genes, DNA methylation and exposures underlying COPD. We were able to identify several (epi-) genetic loci in never-smokers and showed that the effect of air pollution and occupational exposures on lung function is mediated by DNA methylation. Nevertheless, this is only the beginning of understanding the mechanisms playing a role in non-smoking related COPD. I hope others will give more attention to never-smokers in their studies by recognizing the importance of this work for never-smoking COPD patients and possible contributions towards precision medicine.

Study population

In this thesis, we were able to include 5,070 never-smokers in the genetic studies and 1,561 subjects in the methylation studies (1,017 in the air pollution exposure study). This is a relatively large number compared to the few studies so far in exclusively never-smokers or studies assessing methylation-exposure associations. However, it is a small number compared to other population-based GWAS and EWAS studies on for example height, birth weight or cancer. To be able to identify more biological relevant genetic or epigenetic variants related to non-smoking COPD, we need to increase the sample size.
and collaborate in consortia. As we showed in chapter 4 it is however of great importance to make sure that a clear (and the same) COPD definition is used in all studies. We showed that the use of a slightly different airway obstruction definition in the discovery cohort resulted in a different selection of SNPs for replication. COPD encompasses multiple phenotypes, like emphysema, chronic bronchitis or small airways disease, which cannot be distinguished from each other using FEV$_1$, FVC or FEV$_1$/FVC lung function measurements. In addition, there are many risk factors for COPD and various exposures can induce changes in different underlying pathways, possibly leading to the same COPD phenotype. It might thus be beneficial to have a more homogeneous study population rather than increasing the sample size. For example, in our DNA methylation studies we performed the analyses in the complete cohort and stratified by smoking status. In most occasions we were able to identify and replicate more significant methylation sites in never-smokers compared to the more heterogeneous complete cohort. In addition, other differentially methylated sites and SNPs were identified in never-smokers compared to smokers, pointing again to different underlying (epi-)genetic pathways. But how can we define a homogenous study population? Several GWA studies have investigated specifically emphysema or chronic bronchitis, but largely only in smokers. SNPs annotated to HHIP and the CHRNA region were consistently identified in GWA studies on emphysema, and the gene FAM13A was identified in a GWAS on chronic bronchitis. These well-known genes were also identified in our studies and we showed that HHIP and FAM13A are general COPD associated genes, independent of smoking status (Chapter 2 and 3). Maybe we therefore should not study the classical COPD phenotypes, like emphysema and chronic bronchitis, but rather clinical COPD subtypes that have been developed in an unbiased way. Firstly, several studies have used cluster analysis to identify COPD subtypes and identified symptoms and comorbidities to be important classifiers. Secondly, we could investigate more pathology specific outcomes, since for example based on pathology several types of emphysema are being recognized which could be due to different mechanisms. For future studies, it would be useful to have computed tomography scans (CT-scan) to assess the degree and subtype of emphysema and possibly data on symptoms and comorbidities. Subjects could also be subjected to an allergy test to assess if different mechanisms play a role in allergic COPD or to exclude the group from the analysis.

We used the broad COPD definition of FEV$_1$/FVC< 70% in our methylation study and were not able to identify any DNA methylation sites genome-wide significantly associ-
ated with COPD in either never- or current-smokers (Chapter 5). This is in contrast with previous published literature, that identified many very significant CpG sites and we already discussed differences between the studies in chapter 5. It is however important to take into account that we used pre-bronchodilator measurements to define airway obstruction. To exclude the asthmatic phenotype and have a better defined study population, post-bronchodilator measurements should ideally have been used. Furthermore, whole blood was used to generate the methylation data. We know DNA methylation is tissue specific and can have a different function depending on the location of the CpG site to the gene.\textsuperscript{17} Perhaps more pronounced DNA methylation changes are present in specific lung tissue cells, like fibroblasts, bronchial epithelial, alveolar, or endothelial cells. Fortunately, for both air pollution and occupational exposures, we were able to identify genome-wide significant differentially methylated sites and regions in blood (Chapter 7–9). But do the assessed environmental exposures induce methylation changes in lung tissue as well- It would thus be interesting to investigate DNA methylation sites/regions in lung tissue, and in addition assess if the results in lung tissue and blood are similar. Since lung tissue also consists of many different cell types, we could for example assess specifically epithelial cells (first in contact with exposure), fibroblasts or smooth muscle cells.

Prospective longitudinal studies are warranted to assess associations between environmental exposures and DNA methylation changes, and their effect on lung function or COPD. For all the studies described in this thesis, associations have been studied cross-sectionally in the Dutch population and therefor causality could not be assessed. Longitudinal data on lung function and occupational exposure is available for the Vlagtwedde-Vlaardingen cohort, but data on DNA methylation and air pollution are lacking. The second round of follow-up is in progress for LifeLines, and hopefully the cohort will continue collecting data every 5 years. When this data becomes available, efforts should be made to validate the identified methylation regions in these follow-up samples, as well as to assess the direction of the effect. It is important to know if the change in methylation level is due to the exposure or if it is a consequence of a developing disease.

**Replication of results**

To assess if identified associations are true biological findings and not false positives, it is very important to replicate the results in independent cohorts. This is in my opinion
even more important than reaching genome-wide significance. The general accepted genome-wide significant threshold for genetic imputed studies is $5 \times 10^{-8}$, based on the Bonferroni correction for 1 million tests, since it was estimated that there are approximately 1 million independent SNPs in the human genome. In our genetic studies we only used genotyped SNPs and we used a lower cut-off ($10^{-5}$ or $10^{-4}$) to lower the risk of false negative findings. We only considered a result replicated when the effect estimate was in the same direction. This means that in different populations the same effect is found. This is stronger proof of a true biological finding then having appropriate $p$-values, since they only reflect a probability.\textsuperscript{18–20} Unfortunately, we were not able to find suitable replication cohorts for all our studies, especially for the occupational exposure and methylation studies. Most other studies do not have both 450K methylation data and data on job exposures. In addition, in the case this data was available, the number of highly exposed subjects was still insufficient to replicate our results.

**Genetics of COPD in never-smokers**

We identified several SNPs associated with lung function or COPD in never-smokers annotated to known or novel susceptibility genes (Table 1). The genes \textit{HHIP}, \textit{FAM13A}, \textit{FABP7} and \textit{NFYC} are susceptibility genes for the development of airway obstruction and our data suggests that \textit{MECR} plays a role in specifically small airways obstruction. Since with GWA studies we do not measure all SNPs present in the human genome, these identified SNPs are most likely not causal but rather capture a genetic region associated with COPD or lung function. Performing additional analyses using sequencing or imputed genetic data could help narrowing the causal region or identify the causal SNP. To understand the functional consequences, we need to understand how the SNP for example affects gene expression or changes protein structure, and furthermore what the role of this gene is in COPD. We therefore performed eQTL studies to assess the association between SNPs and gene expression levels in lung tissue and found significant association for all SNPs except one (Table 1).
Discussion and future perspectives

Table 1. Identified SNPs within this thesis associated with lung function or COPD in never-smokers.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>MAF</th>
<th>Outcome</th>
<th>Effect estimate</th>
<th>eQTL lung</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1512282</td>
<td>HHIP</td>
<td>41%</td>
<td>FEV₁/FVC (%)</td>
<td>0.60</td>
<td>HHIP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FEF₂₅₋₇₅ (mL/sec)</td>
<td>98 / 72*</td>
<td></td>
</tr>
<tr>
<td>rs6849143</td>
<td>FAM13A</td>
<td>43%</td>
<td>FEV₁/FVC (%)</td>
<td>-0.54</td>
<td>FAM13A/TIGD2</td>
</tr>
<tr>
<td>rs2452785</td>
<td>MECR</td>
<td>20%</td>
<td>FEF₂₅₋₇₅ (mL/sec)</td>
<td>105 / 114*</td>
<td>MECR/EPB41</td>
</tr>
<tr>
<td>rs7519348</td>
<td>NFYC</td>
<td>33%</td>
<td>FEV₁/FVC&lt;70% (OR)</td>
<td>1.36</td>
<td>NFYC-AS1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FEV₁/FVC&lt;LLN (OR)</td>
<td>1.37</td>
<td></td>
</tr>
<tr>
<td>rs6913003</td>
<td>FABP7</td>
<td>5%</td>
<td>FEV₁/FVC&lt;70% (OR)</td>
<td>1.90</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FEV₁/FVC&lt;LLN (OR)</td>
<td>1.83</td>
<td></td>
</tr>
</tbody>
</table>

* In subjects without large airway obstruction

MAF = Minor Allele frequency

Based on family and twin studies, we know that the genetic heritability of lung function and COPD is 20–40 %.²¹⁻²³ Even though many genes have been associated with COPD, only a small percentage of the genetic heritability can be explained by these genes. Therefore small effects of many genes may contribute to COPD. We only identified a few SNPs associated with lung function or airway obstruction in never-smokers, but they are quite common (high minor allele frequencies) and therefore might have an impact on the prevalence of COPD worldwide. In addition, it is important to assess the interaction, or independent and additive effects of the identified genetic variants. We assessed the additive effect by combining the HHIP and FAM13A SNPs into one genetic risk score. We saw a clear additive effect of both genes with a credible clinically relevant effect. Combining information of more genes could give us further insights into the underlying pathways in COPD. Already there are tools available to assess known pathways, but in our studies these did not yield additional information since only very general pathways including many genes were identified. Perhaps pathways have yet to be discovered that underlie non-smoking related COPD.

Often people ask “what is your favorite gene-”. Well, my favorite gene is HHIP since our studies illustrated that HHIP is an important lung developmental gene of both the small and large airways. In more detail, we found that a SNP annotated to HHIP was associated with spirometric measurements of both the large (FEV₁/FVC) and small airways (FEF₂₅₋₇₅), and in addition similar effect estimates were found in never- and ever-smokers. We also showed that the SNP was associated with gene expression levels of HHIP in lung tissue. Another SNP annotated to HHIP was associated with airway obstruction in only ever-smokers, but this association could not be replicated. This data points towards
HHIP being an important lung developmental gene and subjects carrying variants near HHIP might thus suggestively have a lower lung function at birth compared to non-carriers. This may be due to insufficient growth and lung hypoplasia, and deficiency of the hedgehog pathway may have an effect on airway branching or smooth muscle cell differentiation later in life as well.\textsuperscript{24} HHIP might therefore be an interesting drug target, but more research into the specific role of HHIP during lung development and its role later in life is warranted.

Environmental exposure assessment

To date, we are the first to perform a large genome-wide methylation study assessing the association between air pollution or occupational exposures and DNA methylation levels. We identified numerous methylation regions in the genome affected by these exposures and this information could lead to understanding the involved pathways. We estimated the degree of outdoor air pollution exposure based on the subjects home address using land-use regression (LUR) models and estimated occupational exposure levels in the current or last held job using a job exposure matrix (JEM). Nevertheless, the risk of misclassification is significant for both types of exposure estimations. Moreover, air pollution is a mixture of hundreds of pollutants originating from industry, traffic, heating, and other sources. Throughout the day people are exposed to different sources of air pollution since they move from place to place and likely do not stay at home the entire day. Even when they would stay at home all day, they will be exposed to indoor air pollution from cooking, heating, or burning candles. In our studies we were not able to account for this, since we used measured annual mean ambient air pollution levels to estimate the level of NO\textsubscript{2} and PM exposure at the home address (LUR models). Previous studies have shown that outdoor air pollution measurements highly correlate with personal (indoor) air pollution measurements.\textsuperscript{25,26} However, the actual level of air pollution exposure could be different between the two measurements. Improvements on assessing personal air pollution exposure levels in large population based studies are therefore warranted. For example, nowadays many apps and small portable devices have been developed to measure personal exposure. The mobile phone apps can be used to track movement and based on this data more accurate estimations could be made about (personal) air pollution levels. Even better, the small portable devices would give even more accurate information on daily (or even hourly) air pollution and occupational exposure. During the ERS 2016 in London the clean space app and tag (personal air pollution smart sensor) were presented. The app was developed to create
more awareness about air pollution in London and based on given air pollution data users can earn CleanMiles if they make small changes in the way they travel to avoid highly polluted areas in London. The information that the app and tag collect could be of great benefit to research and will not take a lot of effort from the participant. Other examples of small portable sensors are: TZOA (measures air quality, temperature, humidity, atmospheric pressure, ambient light and UV exposure), AirBeam (uses light to measure PM2.5), and the Clarity device (low cost device that periodically samples surrounding air to identify pollutant particles) (Figure 1).27

Figure 1. Small portable devices to measure air pollution.

Studies have shown that using a job exposure matrix is a reliable method to estimate occupational exposures. There is a high correlation with expert opinions and it performs better in estimating job exposures compared to exposure self-reports.28,29 Of course there is still a possibility of misclassification, also considering that occupational exposures are in addition dependent on the type of task, the taken protective measures and used ventilation. For this reason a task exposure matrix (TEM) was developed, and this could possibly reduce the exposure misclassification.30 The previous mentioned portable measuring devices could also be used to assess personal occupational exposure levels, assuming no mask is being worn. Moreover, using a JEM does not allow to assess specific chemical compounds present in the occupational exposure and different types of jobs classified into the same exposure category might contain different chemical compounds as well. In addition, we saw that within one job, exposure to several types of occupational exposure can occur, like gases/fumes, mineral or biological dust at the same time. This makes it challenging to disentangle the effects of specific occupational exposures. Indeed we saw that associated methylation sites were overlapping between exposures. This could indicate a general mechanism, like oxidative stress, or could just
be due to co-exposure. We did not additionally adjust our studies for co-exposure to other occupational exposures, so the answer remains unclear.

**Environmental exposures and lung function**

Interestingly, our group found previously that biological dust is most strongly associated with small airways obstruction and not with large airway obstruction.\(^{31,32}\) Biological exposures may include small particles of spores, pollen, viruses, cotton, wood, flour, textile and paper, which may reach deep into the lungs.\(^{33,34}\) In addition, joint exposure to both smoking and occupational exposure (VGDF) was found to increase the risk of COPD even more. This information might explain why for biological dust most significant differential methylation regions (DMRs) were identified in current-smokers. In contrast, for gases/fumes and mineral dust most significant DMRs were identified in never-smokers. The particle size of the pollutant could play a significant role. In general, particles with a size $> 10$ microns are filtered out by cilia and mucus in the throat or nose, smaller particles ($10$–$0.1$ microns) are respirable and may penetrate into the bronchi, bronchioles or alveoli, and particles $< 0.1$ microns might even pass through the lungs into the blood.\(^{35}\) Interestingly, we see in figure 2 that both viruses (biological dust) and tobacco smoke are smaller in size compared to ground talc (mineral dust) or metallurgical dust and fumes. In addition, both spores and pollen are $> 10$ microns and thus probably do not reach the lung. Biological dust from viruses might therefore penetrate deeper into the lungs and play a role in specifically small airway obstruction compared to the other occupational exposures which were also associated with large airway obstruction.\(^{31}\)

Noteworthy, overlapping DNA methylation regions were found between gases/fumes, mineral dust and pesticide exposure, but not with biological dust.
Discussion and future perspectives

In chapter 6 we found a positive association between FEV$_1$/FVC and air pollution due to a small non-significant effect on FEV$_1$ and a larger significant effect on FVC within LifeLines ($n = 51,855$). A similar observation was made in chapter 7 where we assessed if DNA methylation mediated the association between air pollution and lung function in a subset of Lifelines ($n = 1,017$). In addition, also in chapter 8 we found a positive association of occupational exposures with FEV$_1$/FVC and FEF$_{25-75}$ levels in the same subset of LifeLines. In previous research of our group, however, we found in the complete LifeLines cohort a negative association of occupational exposures with FEV$_1$, FEV$_1$/FVC and airway obstruction ($n = 11,851$). This opposite direction of effect could be due to the non-random selection of the LifeLines cohort for our methylation studies to be able to investigate the largest possible contrast between exposures. We selected relatively equal numbers of subjects with and without COPD, cigarette smoke exposure and high levels of job related exposures. Therefore, ex-smokers and subjects with low exposure to the composite job exposure measurement VGDF (vapors, gases, dusts, or fumes) were not included. Our methylation study data thus does not reflect the general population, but does allow for assessing mediation of DNA methylation between exposures and lung function within the same person.
Environmental exposures and DNA methylation

Similar as for SNPs, we need to further understand the functional consequences of the identified methylation regions. We started by assessing the association between the methylation sites and gene expression levels, and in addition assessed if the CpGs mediated the association between an environmental exposure and lung function levels. For many CpGs we found associations with gene expression and in addition some were also significant mediators. This information already points towards possible mechanisms, but it is essential to pinpoint the causal loci to give them a clinical purpose. For example, is a single methylation site responsible for the difference in gene expression or are multiple sites- The position of the CpG site relative to the gene determines its possible function. For instance, CpG sites in the transcription start site can silence genes and cause long-term repression, CpGs in the gene body can stimulate transcript elongation and regulate splicing, and CpGs in repetitive regions assist in chromosomal stability or suppress transposable elements, or they can induce cytosine (C) to thymine (T) mutations. Having higher methylation levels can thus both be detrimental (silence functionally needed genes or induce mutations) or beneficial (stimulate transcription or suppress transposable elements). Already, some methylation inhibition drugs are used in clinical practice, like 5-Azacytidine which globally reduces DNA methylation and thereby may activate tumor suppressor genes. Currently, this drug is used to treat myelodysplastic syndromes like acute myeloid leukemia. Since this drug is nonspecific and the role of DNA methylation in COPD is poorly understood, it is not used as a treatment in COPD. Successful efforts have been made using the genome editing tool CRISPR-Cas9 to make the demethylation-action site specific. The CRISPR-Cas9 system can be used to add or alter specific sections of the DNA. A specific guide RNA binds to the DNA and recruits Cas9 which cuts both strands of DNA at the bound location. Next, DNA repair mechanisms are mobilized to the location and during this stage specific DNA editing can take place. To date, a clinical trial has been started to use genetically edited cells by CRISPR-Cas9 to treat various types of cancers. Epigenetically edited cells by CRISPR-Cas9 as treatment have not been used in clinical settings yet. Perhaps our identified sites could be therapeutic targets for COPD or reverse changes induced by exposures, because “voorkomen is beter dan genezen” (to prevent is better than to cure).

In this thesis we only focused on one epigenetic mechanism, DNA methylation. However, histone modifications or RNA-based silencing are also important epigenetic mecha-
Discussion and future perspectives

Mechanisms which can induce changes in gene expression as well. Yet, these mechanisms are technically more difficult to study and therefore currently less investigated.

Integrating omics data

A challenge for future studies will be combining multiple layers of omics data, including (epi)genomics, transcriptomics, proteomics, and metabolomics, to get a complete picture of all mechanisms involved (Figure 3). Especially in complex diseases, like COPD, integrating omics data could aid in identifying novel pathways or improve our understanding of pathways, assessing the interplay between layers and help capture the heterogeneity of the disease. By identifying key factors underlying COPD, we might be able to design better treatments or even cure COPD.

![Figure 3. The “omics” cascade - From genome via transcriptome, proteome and metabolome to phenotype (Adapted from García-Sevillano et al.45).](image)

Two approaches can be taken to combine the data of all the omic-platforms, we can either use a stepwise approach (multi-stage analysis) or perform a meta-dimensional analysis. For the latter option, multi-layer network-based frameworks and multilevel Bayesian models have been developed, but these involve complex computations. To date, the most commonly used method is the three-stage regression method. First, SNPs are associated with an outcome, next QTL analyses are performed (associations between the SNP and gene-expression (eQTL), DNA-methylation (mQTL), protein levels (pQTL) or metabolites), and subsequently the results of the QTL analyses are correlated with the outcome of interest. This is similar to the analyses described within this thesis, but we only used epi(genetic) and transcription data from different studies and subjects. Ideally, we would need omics data from all layers to be from the same well characterized subjects (with exposure and COPD related data) and also from the same cell type (possibly using single-cell technologies). This can be very costly and time consuming, but as a start we could make use of publicly available datasets from for example the Encyclopedia of DNA Elements (ENCODE) or The Cancer Genome Atlas (TCGA) projects. Data on exposures and lung function are however not available for these data sets, and
in addition the data do not originate from the same subjects. Yet, it would give us the opportunity to assess the associations between omics layers and get a better picture of all mechanisms involved. My advice for future research would be to invest in getting data on all omics layers by expanding existing databases like the lung tissue database. To start integrating the information, we can first use multi-stage analyses, but time should be invested to find the optimal multi-layer approach. This approach should lead to better insights into the underlying mechanisms in COPD.

Never-smoking female COPD patients
Finally, it is intriguing that within the group of patients with COPD in the Netherlands, the proportion of females that has never smoked is higher than the proportion of males that never smoked (27% versus 7%). It was suggested that this among others could partly be due to sex hormones and early menopause has been associated with increased risk of multiple diseases. In addition, being in menopause was associated with lower lung function and more respiratory symptoms. It would therefore be interesting to investigate the influence of age at menopause and hormonal status on lung function decline and the development of COPD. This is one of the objectives within the Horizon 2020 EU-funded Aging Lungs in European Cohorts (ALEC) project, of which I will be a part of.

CONCLUSION

In the current thesis we have shown that genes, DNA methylation and environmental exposures play a role in COPD in never-smokers. Several susceptibility genes for large or small airway obstruction in specifically never-smokers were identified and altered DNA methylation levels upon air pollution and occupational exposures could be the way how these exposures affect lung function. The identified genetic and epigenetic loci could potentially have a clinical purpose, but more research into their function is warranted and several improvements in data collection can be made. For future research, I would suggest to focus on two types of studies. First, we should try to integrate multiple layers of “omics” data to understand the complex mechanisms involved in the COPD phenotypes. Second, we should study the effects of the environmental exposures on epigenetics and lung function longitudinally.
References


