General discussion

In this thesis the genetic and non-genetic determinants and clinical consequences of impaired lung function were studied. Especially, genetic epidemiological studies had a great share in this work. Here, comprehensive study designs were used including GWAS, EWAS, and Genome-wide linkage to investigate which genetic and epigenetic changes are significant contributors to disease onset and progression. For this, we integrated population genetics with experimental clinical research. In this chapter, the main findings, research challenges, clinical relevance and future directions are discussed.

MAIN FINDINGS AND CHALLENGES

Epidemiology of COPD

Epidemiological research is crucial to gain insight into the occurrence of diseases, to support informed decision making and develop public health interventions and strategies to address risk factors (1). Alongside with other research disciplines, epidemiological research is essential and aims to answer questions like: how big is the problem? What is the risk of acquiring the disease in a certain amount of time? And which risk factors are important?

In **chapter 2**, we studied the epidemiology of COPD. COPD is the third leading cause of death worldwide and thereby a major public health problem (2). In de last decade, the number of COPD cases worldwide increased by 28% and this number is expected to grow in the future (3). We investigated the epidemiology of COPD in 15,000 participants of the Rotterdam Study who were followed for up to 25 years. In the absence of a clinical diagnosis of asthma in the electronic medical records, COPD was diagnosed based on a pre-bronchodilator obstructive spirometry (FEV_1/FVC < 0.70). In absence of an interpretable spirometry within the Rotterdam Study, cases were defined as having COPD diagnosed by a physician on the basis of clinical presentation and obstructive lung function measured by the general practitioner or respiratory physician. We reported on the prevalence and the incidence of the disease, and investigated differences by sex and smoking status.

One of the challenges in epidemiological cohorts studies is, that researchers often deal with historical changes that they have to consider in the sake of a good interpretation of the study results. In the Rotterdam study for example, spirometry was introduced after 2001, meaning that for a subset of COPD cases the diagnosis was based on medical chart records only. Should we have presented the prevalence and incidence data of COPD in the total population only, then we would have missed an interesting piece of information that enriches our knowledge about the incidence of COPD in different settings. Subgroup analyses were required to highlight different study

results (**Table 1**). A higher incidence of COPD was found when diagnosis was based on spirometry (only) as compared to a clinically based diagnosis, which can be attributed to the inclusion of asymptomatic COPD subjects who frequently still have mild airflow limitation. Since mild COPD cases rarely seek medical attention, the incidence of COPD is frequently underestimated in clinical settings.

Table 1: Prevalence and incidence data according to different classification methods in the total cohort and in the sub-groups (spirometry versus medical charts group)

	Spirometry data	Medical records data	Combined data
	N=7,153	N=7,466	N=14,619
COPD GOLD			
Prevalence	5.3%	4.2%	4.7%
Incidence	11.7/1,000PY	5.8/1,000PY	8.9/1,000PY
COPD LLN			
Prevalence	3.4%	4.2%	3.8%
Incidence	5.2/1,000PY	5.8/1,000PY	5.5/1,000PY

COPD: Chronic obstructive pulmonary disease; GOLD: Global initiative for Obstructive Lung Disease; LLN: lower limit of normal; PY: person years

Another historical change is the introduction of "better" (i.e. more accurate) cut-off values of FEV_1/FVC to define airflow limitation. In the same chapter, we presented the incidence figures of COPD according to either the fixed 0.70 ratio of FEV_1/FVC (as recommended by the Global initiative for Obstructive Lung Disease [GOLD]) or the lower limit of normal (LLN; as recommended by the Global Lung function Initiative [GLI]) (**Table 1**) (4-7). The question remains whether we really have to replace the fixed ratio with the LLN. Several studies claim that using the fixed 0.70 FEV_1/FVC ratio overestimates the diagnosis of COPD in older individuals due to the inclusion of asymptomatic subjects with mild COPD (GOLD I) (6, 8, 9), while others state that mild COPD might be a good marker for preclinical disease (6). Importantly, the new criteria does not consider subjects with mild COPD (GOLD I) as "true" COPD cases, even though we have shown in the Rotterdam Study that even subjects with mild airflow limitation (i.e. mild COPD) have an increased risk of cardiovascular mortality (10), all-cause mortality (11) (**Figure 1**), and future severe exacerbations compared with control subjects with normal spirometry (12).
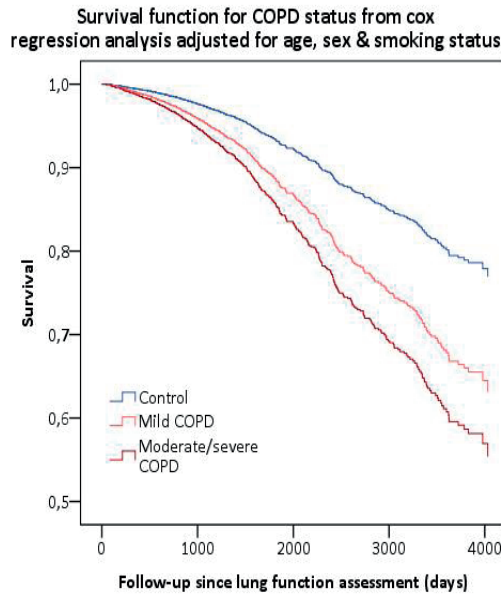
Survival function for COPD status from cox regression analysis adjusted for age, sex & smoking status

Figure 1 All-cause mortality in COPD in The Rotterdam Study, stratified by COPD severity

The global initiative for chronic obstructive lung diseases (GOLD) defined COPD as a preventable and treatable chronic respiratory disease (13, 14). Although research efforts broadened our understanding of the disease, we are still unable to unravel the exact pathophysiological mechanisms of this disease, nor we are able to define accurate biomarkers of COPD (15-21). Thereby, we are left with a progressive disease that forms a huge burden on the quality of life of millions of people worldwide (14). The only bright spot for patients with COPD are medications for symptoms relief that help them improve their quality of life (22). However, this change in COPD treatment forms a methodological challenge in epidemiological research of the 'natural course' of COPD, because the more treatment options are available, the more difficult it is to capture the differences in severity that arise, due to effect of treatment.

Finally, COPD is an age-related disease. Patients with COPD often present with multi-morbidity including cardiovascular and cerebrovascular disease (22). For this very reason, researchers encounter difficulties studying associations between COPD and other age related diseases, such as dementia, because the more advanced the disease stage is, the more researchers will deal with informative missingness in their data, leading to a possible bias in the generated results (23).

Genome-wide association studies (GWAS)

GWAS associations can typically tag multiple genetic variants in the same region that are highly correlated (or in other words, in linkage disequilibrium (LD)). Correlated

variants have an equal likelihood to be causally associated with the disease of interest, when assuming no independent effects. The idea behind LD relates to the non-random association between one or more loci on the genome (24). The LD between SNPs in the same chromosomal region is calculated based on information of the observed haplotype frequency of the two SNPs of interest and the frequency of each individual SNP (24-26). Although there are many ways proposed to calculate LD (27, 28), it is nowadays typically expressed as a measure of squared correlation (r^2), or as the relative measure of disequilibrium compared to its maximum, as proposed by Lewontin , the D prime (D') (26). However, many web based applications, that provide summaries and visualizations of GWAS results often base their LD calculations on r^2 only. It is important to realize that r^2 can only reach its maximum if the allele frequency of the two SNPs are similar and the two haplotype frequencies are bigger than zero (25, 29). While D' can be large if any allele of the first locus is seen with one allele of the other, and thereby does not depend on allele frequencies as it is the case with r^2. As we are now in the era of conducting GWA studies with better imputation of rare variants, it is desirable to realize that the use of r^2 might lead to misleading results about the true extent of the LD, simply because of its properties. D' might be more useful in case we are interested in LD measures using a less frequent variant. Nevertheless caution is warranted in the interpretation of D' as well, since D' can be inflated when the sample size is small or if one of the two alleles is extremely rare (29, 30).

The statistical power to detect associations in GWAS depends on several factors such as sample size, effect size of the genetic variant(s), the heterogeneity of the phenotype of interest, measurement errors, the distribution of the causal variants in the studied population, their frequency and the correlation (LD) with the observed genotyped variants (31, 32).

In **chapter 3**, we presented two GWA studies of lung function. In the first GWA study we investigated a multi-ethnic cohort with more than 90,000 individuals from 22 studies, of whom 60,000 of European descent. Here, we calculated the heritability and genetic determinants of FEV_1, FVC and FEV_1/FVC, and identified over 50 novel loci, of which sixteen genes encode proteins with predicted or established drug targets (33). In the second GWA study, we calculated the heritability of and studied the genome-wide associations with diffusing capacity of carbon monoxide in approximately 8,000 individuals. Here, we found that the *ADGRG6*, locus was significantly associated with DLCO/VA and that the expression of this gene was decreased in patients with COPD and in subjects with decreased diffusing capacity (DLCO/VA) (34).

Sample size
The two GWA studies show roughly similar heritability estimates but differ considerably in the number of identified loci (50 novel loci versus 1). This example is a very nice

illustration of how sample size (60,000 versus ~8,000 individuals) can increase the power to detect (novel) genome-wide significant associations.

Sample size is one of the most important factors that influence the power to find an association in GWA studies. However, since we are dealing with age related traits of lung function, the question remains whether it is justified to add individuals from all age groups in order to increase sample size and thereby gain power, or to restrict the study population to specific age groups (e.g. older individuals) in order to decrease trait variability and thereby to increase the power.

Heterogeneity

Phenotypic heterogeneity reduces the power to detect statistically significant associations and decreases the magnitude of the risk estimates that are ascribed to genetic variations (35). In the GWAS of diffusing capacity of carbon monoxide, we had a relatively small sample size to begin with, however we tried to gain more power by reducing the heterogeneity of the trait by using not only interpretable but also reproducible measurements in the analyses, so that only high quality lung function measurements were included (32, 35).

Intuitively, the greatest power might be obtained when these two important factors (sample size and heterogeneity) are in balance. Manchia and colleagues suggested that accurate phenotype delineation is maybe even more critical for new genetic discoveries than increasing the sample size (35).

Genetic correlation

So far, GWAS had an important role in genetic discoveries beyond detecting trait-SNP associations (31). GWAS summary statistics can be used in Mendelian randomisation studies to test causal relationships, or to quantify the genomic architecture by assessing the LD structure. These can also be used to detect and quantify pleiotropy by means of polygenic risk scores or genetic correlation estimations. In **chapters 3.1** and **3.2**, we used summary statistics of the GWASs to test genetic correlations between lung function and other traits. An important challenge in genetic correlation studies is, that they are sample size dependent (36, 37). In other words, lack of statistically significant correlation does not mean lack of pleiotropy, but could be due to lack of power to detect pleiotropy. Another challenge in genetic correlations studies is, that genetic correlation estimates might be counter intuitive, for example in the case of a negative genetic correlation between two lung function measures; FEV_1 and DLCO/VA (as measured using data from the Rotterdam Study: -0.35 (SE 0.13), p-value=0.009). Several reasons are possible to explain negative genetic correlations (**Figure 2**), these include: A) direct effect of the causal SNP on the trait, where the SNP has a positive effect on one trait and a negative effect on the other, B) a causal SNP that has a positive effect on the first

trait, which might be in LD with another causal SNP that has a negative effect on the second trait, C) a causal SNP that influences the traits via a mediator that has a positive effect on the first trait and a negative one on the second, and finally as an alternative for explanation C, D) a causal SNP that affects the first trait which in turn affects the second trait in an opposite manner. It is not straightforward to elucidate the exact reason of a negative genetic correlation based on the GWAS results only. Therefore, many methods are now used to detect genome-wide, regional or single-variant pleiotropy (37-39). These methods provide a quantitative assessment of pleiotropy, and have an important role in understanding shared genetic effects.
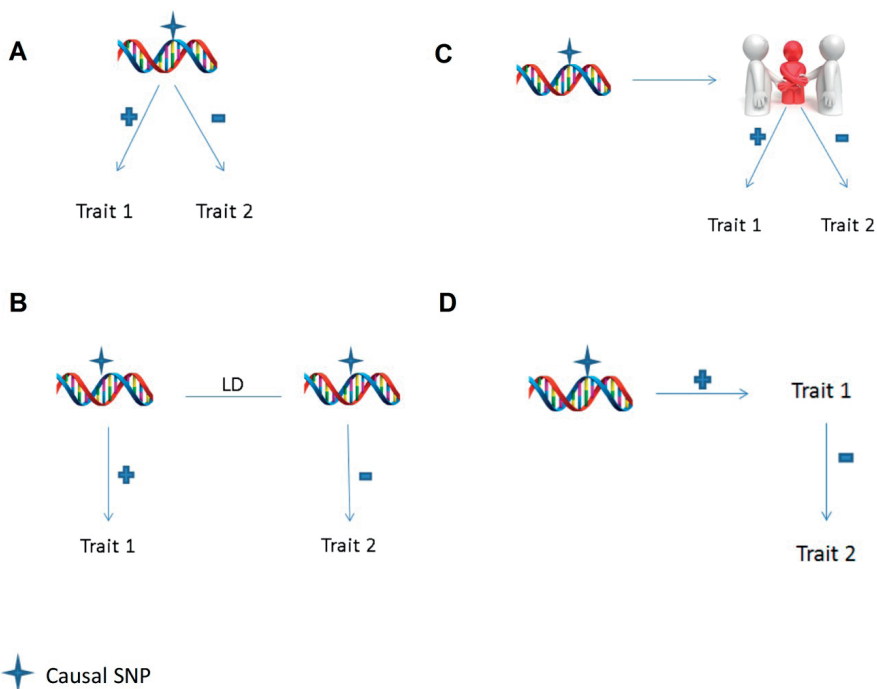

Figure 2 Interpretation of a negative genetic correlation between traits with positive phenotypic correlation

Epigenome-wide association studies (EWAS)

In **chapters 3.3** and **3.4**, we studied epigenetic associations of blood DNA methylation with occupational exposures and diffusing capacity, respectively. Analyses of the first study (**chapter 3.3**) were restricted to 903 never-smokers from **LifeLines**. The association between occupational exposure to gases/fumes, mineral or biological dust and blood DNA methylation was assessed. Several associations were identified and the results were validated in the Rotterdam Study. Finally, methylation-expression associations were assessed. In the second study (**chapter 3.4**), the association between

DLCO and DLCO/VA and blood DNA methylation was assessed, after adjustment for several covariates including smoking. Data from the Rotterdam Study was used for discovery and data from the Framingham Heart Study served as replication panel. Two methylation sites were identified and replicated and the role of smoking related genes on lung function was discussed. In both studies we hypothesised that DNA methylation might act as an intermediate between occupational exposures (**chapter 3.3**) or smoking (**chapter 3.4**) and lung function impairment.

A considerable challenge in EWAS is, that EWAS associations can be causal as well as consequential, implicating major challenges in study design and interpretations of the findings (40, 41). GWAS and EWAS findings are hypothesis generating and need replication in independent population-based or clinic-based cohorts. Replicated findings are interesting candidates for further experimental research.

In the **EWAS** of diffusing capacity of carbon monoxide (DLCO), we hypothesised that the causal effect of smoking would affect methylation levels of the CpG site in the **AHRR** gene. The hypomethylation of the *AHRR* gene is associated with the low expression of another gene, **EXOC3**. We were able to associate the down regulation of mRNA expression of this gene with decreased diffusing capacity in human lung tissue (**Figure 3**). It is worthwhile to mention that although we were able to demonstrate clinical relevance by linking the methylation status to gene expression and consequently lung function decline, the proposed hypothesis is the result of conclusions based on multiple steps where each individual step in the triangle is prone to errors that might lead to biased results and conclusions. For example, residual confounding by smoking might lead to bias in any direction. Also, the biological samples to perform the experimental work on, are often derived from a smaller sample of selected clinical patients (e.g. undergoing thoracic surgery for lung tumours). Therefore caution must be warranted when dealing with such datasets and a well-designed analysis plan is encouraged to avoid bias as much as possible.
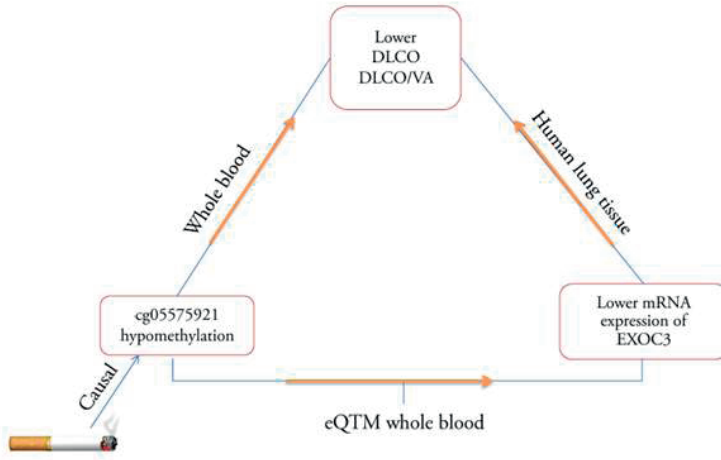
Figure 3 Possible pathway through which hypomethylation of cg05575921 (*AHRR gene)* affects diffusing capacity.

Genome-wide linkage scan

Although larger studies and more accurate data are needed to increase the power for discovery in GWAS, it is unclear whether the findings will help to explain the biggest proportion of the missing heritability. Therefore rare variant analyses, including genome-wide linkage scans might be interesting, as they can be informative for discoveries beyond the ones revealed by GWAS so far (**Table 2**).

Table 2 Possible sources of genetic heritability that might explain the missing heritability problem (42, 43)

Sources of heritability	Causes of missing heritability
Common genetic variants	- Lack of power in GWAS
Rare genetic variants	- Lack of power in WES and WGS - Rare variants might explain only a small part of the heritability - Underrepresentation of rare variants in SNP arrays
Epigenetic variants	- Difficulty to control for epigenetic changes that contribute to variable gene expression
Interaction and epistasis	- GWAS assume additive effects without interaction - Twin studies do take interaction and epistasis into account
Copy number variations, insertions, deletions, inversions	- These variants are largely under investigated - Large sample sizes are needed - Often considered as unapproachable by the current genotyping and sequencing technologies.

GWAS: Genome-wide association study; WES: Whole exome sequencing; WGS: Whole genome sequencing

COPD is a complex disease with genetics contributing to a substantial part of its variability. In **chapter 3.5**, a genome-wide **linkage** scan was performed to discover rare genetic variants underlying COPD in 142 cases clustered in 27 families from an isolated population in The Netherlands; The Erasmus Rucphen Family study (**ERF**) (44). Several linkage peaks were detected (**Figure 4**). Potential causal variants were selected for further association testing with COPD in ERF and the Rotterdam Study. The strengths of this study are that it provides the opportunity to investigators to point-out genetically interesting regions that are associated with COPD. However, a major limitation is often the lack of sequencing information in the affected family members, implicating that no further analysis could be performed (yet) with two of the three discovered linkage peaks.
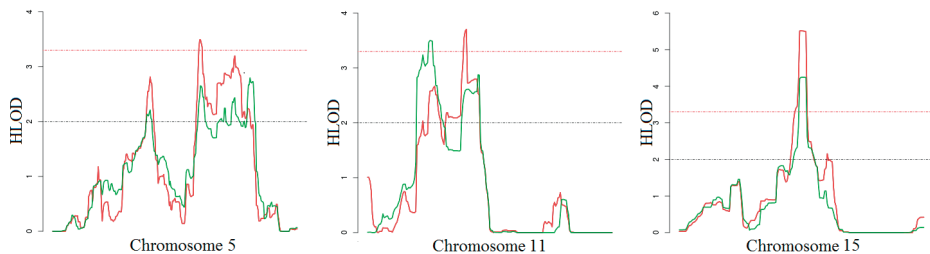

Figure 4 Heterogeneity logarithm of the odds (HLOD) score plot for the chromosomes 5, 11 and 15. X-axis shows chromosome positions and Y-axis shows HLOD score. Red line represents HLOD scores for recessive model and green line for dominant model. Dashed red line represents the level of significance (HLOD=3).

Consequences of impaired lung function

In the fourth chapter of this thesis, we tried to understand the link between impaired lung function or COPD and other comorbidities, including pulmonary hypertension and peripheral arterial disease (PAD).

Pulmonary hypertension

Pulmonary Hypertension is a serious complication in patients with COPD. It evolves due to hypoxic vasoconstriction of the small pulmonary arteries, inflammation, mechanical stress of hyperinflated lungs and loss of pulmonary capillary bed due to emphysema. Pulmonary hypertension may lead to structural changes in the pulmonary arteries and can eventually lead to right ventricular hypertrophy and heart failure (45-47).

Pulmonary artery to aorta (**PA:A**) ratio, defined as the ratio between the diameters of the pulmonary artery and the aorta on transversal images of CT-scans of the thorax, has been shown to be a good surrogate marker of pulmonary hypertension in patients with COPD and a predictor for an increased risk of severe COPD exacerbations requiring hospitalisation (48). In 2,197 participants from the Rotterdam study, we assessed the

Erasmus University Rotterdam

association between 1-SD increased of PA:A and mortality in the general population and in individuals with COPD (**chapter 4.1**). We demonstrated that higher PA:A ratio -as a proxy for pulmonary hypertension- has predictive value on **mortality** in individuals with **COPD**, especially in individuals with low **diffusing capacity**. Importantly, even though an increased PA:A ratio is not completely indicative of pulmonary hypertension (49, 50), the PA:A ratio might be a good proxy for the impact of pulmonary hypertension related mechanisms. Other mechanisms for an increased PA:A ratio include left heart disease, pulmonary inflammation, airway remodelling and vasoconstriction, loss of capillary beds in the pulmonary circulation and hyperinflation. In conclusion, this imaging-based marker - the pulmonary artery to aorta (**PA:A**) ratio - has possibly important clinical utility, as it might help in identifying those COPD patients who are at greater risk of severe exacerbations, hospitalisations and mortality, and thereby provide guidance for a more targeted therapeutic decision making (51).

Peripheral artery disease

Patients with COPD present with multi-morbidity of which most prominently vascular disease e.g. atherosclerosis. Peripheral arterial disease (PAD) is one of the manifestations of atherosclerosis and appears to be often asymptomatic in individuals with COPD. In **chapter 4.2**, we aimed to investigate the association between COPD and future PAD (Ankle Brachial Index ≤ 0.9) longitudinally, and to elucidate the effect of PAD on mortality in individuals with COPD. Therefore , we included 3,123 participants from the Rotterdam study. We found that, individuals with COPD have a higher (almost doubled) risk of developing PAD. People suffering from both COPD and PAD had higher mortality rates compared to people without both diseases. Although the increased mortality could not be directly attributed to PAD as cause, the occurrence of PAD frequently heralds vascular disease in other vascular beds resulting in cardiovascular disease (e.g. ischemic heart disease) and cerebrovascular disease (e.g. stroke).

Future directions

In this section, I discuss my ideas about future research in the field of pulmonary and genetic epidemiology.

Where do we stand now?
Today, we are dealing with two opposite movements in science. On the one hand, the increasing interest of scientific journals to support the open data policies. On the other hand, the General Data Protection Regulation (GDPR), that came into force on the 25th of May 2018, that forces attention and action for personal data protection (52-57). At first sight, open data policy and more data protection seem to oppose each other. There are several examples in the scientific field, where data sharing has been practiced

successfully under the umbrella of ethical and legal norms to the individual privacy. For example, the Global Alliance for Genomics and Health (GA4GH), established in 2013, aims to catalyse genomic science by developing the right legal frameworks to assure responsible global sharing of clinical and genetic data (58-62). Such frameworks are needed, since they offer legal dimensions to research and data sharing between the involved institutes and governments, to protect responsible data sharing. Whether GDPR will prohibit data sharing within and between countries is yet to be discovered; several issues in GDPR are subject for own interpretation and need to be clarified first (54). The good news is that these uncertainties will definitely lead to global discussion about future data sharing in research.

COPD research

Despite the enormous effort that is done so far to understand the causes of COPD, comprehensive knowledge of the disease pathophysiology of COPD is still lacking and epidemiological research in older COPD individuals with more advanced disease is challenging. Early identification of individuals who will develop COPD, will not only help with preventing disease onset or progression, it would also solve important challenges which epidemiological COPD research faces, due to the effect of aging and disease severity. Therefore, future research must focus on understanding disease pathophysiology and finding a suitable biomarker for COPD. For this, results of large (epi)genome-wide association studies must also be considered, as these provide simultaneous information on important disease pathways that are involved in the onset and/or progression of this heterogeneous and complex disease.

Genetic discoveries

In the last decade, many consortia of genetic association studies have been set up (63-65), in order to facilitate the possibility of meta-analysing genetic data across many population-based and clinic based cohorts. We witnessed an increase in genetic discoveries by increasing sample sizes of those studies (up to 3 million in the GIANT consortium with Height and BMI phenotypes). It is likely to assume that this trend of cohorts with increasing sample sizes will continue and that it will probably cumulatively explain a substantial amount of the missing heritability in lung function traits and respiratory diseases. Importantly, these large meta-analysis cannot be well performed without a well-designed analysis plan supported by guidelines for observational studies (66, 67), especially harmonization of the data across the cohorts is needed in order to obtain meaningful and unbiased results.

Another approach to increase the chance of discovery, is to use either better imputation reference panels, such as the haplotype reference consortium (HRC) panel, or to use the more expensive deeper density coverage of variation in the genome using

whole exome sequencing (WES) or whole genome sequencing (WGS) techniques. In this thesis, we present two GWAS in which we used 1,000 genomes as imputation reference panel. Even though we are now in the era of using WGS data with better density of coverage of variation in the genome and larger MAF spectrum, the associated costs are still a hurdle. Therefore, it might be a reasonable next step to use the HRC imputation panel in future GWAS of lung function, given its availability, low costs and reasonable accurate imputations of variants with frequency of 0.1% and higher (68).

While there is a long way ahead of us to find out what we do not yet know in lung function genetics (the so called missing heritability; and whether the discovered genetic variants are causal), there is also a huge need for translational research to elucidate how the – presumably causal - genetic variants contribute to the disease pathogenesis. Previous GWAS findings are excellent candidates for experimental studies to provide insight in the pathophysiology of disease, including knockout or knockdown animal models, experimental studies in human (lung) tissue or studies using relatively novel gene editing techniques, such as the CRISPR-cas9 technique (69). These studies are essential to understand the pathophysiology of the genetic findings and shows the path that we need to follow to link genetic discovery with clinical relevance, and most importantly the identification and validation of novel therapeutic targets (**Figure 5**) (70).
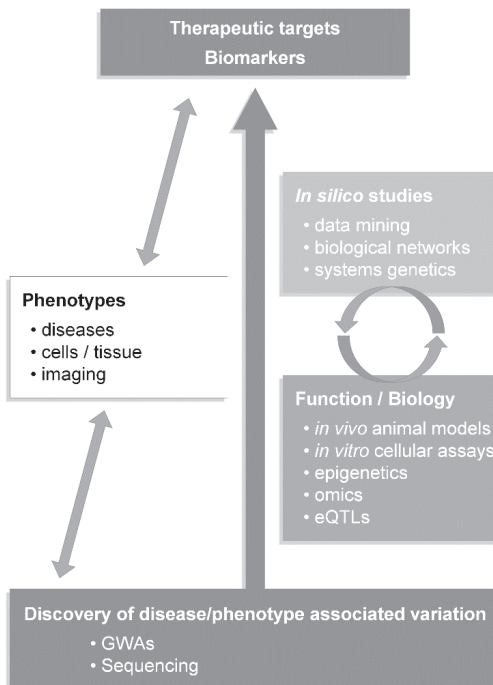

Figure 5 The journey from genetic discoveries based on genetic association studies to finding therapeutic targets and biomarkers for a particular trait or disease.

Clinical observational studies

Regarding the clinical epidemiological studies, our findings could have important implications for disease management in patients with COPD. Indeed, routine measurement of Pulmonary artery to aorta (PA:A) ratio on the chest CT scan (**chapter 3.1**) and/or Ankle Brachial Index screening (**chapter 3.2**) might contribute to timely diagnosis of vascular disease and its consequences.

Implementing PA:A ratio as a prognostic marker for mortality in patients with moderate-to-severe COPD might help with risk stratification and better disease management. There is a significant increase in using chest CT imaging in clinical practice, either in the context of screening for lung cancer in (heavy) smokers or for establishing a diagnosis of pulmonary embolism or ischemic heart disease. On these (screening or clinical) CT scans, also the PA:A ratio should be measured systematically since it provides prognostic information, especially in patients with COPD. Still, several questions remain unanswered. It is interesting to know whether PA:A ratio also applies for patients with mild COPD. Finally, repeated measurement of PA:A ratio in the population could be useful to understand trajectories of PA:A change and their accompanied effect on mortality in COPD.

There is increasing evidence that COPD and peripheral artery disease (PAD) are associated, and our study suggests that COPD might contribute to the development of PAD. However, in a next step it would be interesting to investigate the bi-directionality of the association between COPD and PAD, and specifically to investigate the vascular hypothesis in the development of COPD and emphysema (71). It is indeed tempting to speculate that not only pulmonary and systemic inflammation, but also alterations in the lung circulation including pulmonary vessel remodelling and endothelial cell death are involved in the pathophysiology of COPD (72).

CONCLUSION

In conclusion, in this thesis we presented our work on the epidemiology of impaired lung function and its consequences on morbidity and mortality. An integrated approach to the study of genetic and non-genetic determinants of impaired lung function is necessary in order to elucidate the pathophysiology of COPD and other respiratory diseases.

REFERENCES

1. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Measures of effect in epidemiological research. Nephron Clin Pract. 2010;115(2):c91-3.
2. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. 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(9859):2095-128.
3. Disease GBD, Injury I, Prevalence C. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017;390(10100):1211-59.
4. Terzikhan N, Verhamme KM, Hofman A, Stricker BH, Brusselle GG, Lahousse L. Prevalence and incidence of COPD in smokers and non-smokers: the Rotterdam Study. Eur J Epidemiol. 2016;31(8):785-92.
5. Housset B. [The lower limit of normal: toward a new COPD detection strategy?] La limite inferieure de la normale : vers une nouvelle strategie de detection de la BPCO ? Rev Mal Respir. 2010;27(9):991-2.
6. Mohamed Hoesein FA, Zanen P, Lammers JW. Lower limit of normal or FEV1/FVC < 0.70 in diagnosing COPD: an evidence-based review. Respir Med. 2011;105(6):907-15.
7. van Dijk WD, Gupta N, Tan WC, Bourbeau J. Clinical relevance of diagnosing COPD by fixed ratio or lower limit of normal: a systematic review. COPD. 2014;11(1):113-20.
8. van Dijk W, Tan W, Li P, Guo B, Li S, Benedetti A, et al. Clinical relevance of fixed ratio vs lower limit of normal of FEV1/FVC in COPD: patient-reported outcomes from the CanCOLD cohort. Ann Fam Med. 2015;13(1):41-8.
9. Izquierdo Alonso JL, De Lucas Ramos P, Rodriguez Glez-Moro JM, grupo de estudio C. The use of the lower limit of normal as a criterion for COPD excludes patients with increased morbidity and high consumption of health-care resources. Arch Bronconeumol. 2012;48(7):223-8.
10. Mattila T, Vasankari T, Kanervisto M, Laitinen T, Impivaara O, Rissanen H, et al. Association between all-cause and cause-specific mortality and the GOLD stages 1-4: A 30-year follow-up among Finnish adults. Resp Med. 2015;109(8):1012-8.
11. Odeyemi YE, Lewis O, Ngwa J, Dodd K, Gillum RF, Mehari A. Does Low FEV1 in Addition to Fixed Ratio and/or Lower Limit of Normal of FEV1/FVC Improve Prediction of Mortality in COPD? The NHANES-III-linked-mortality Cohort. J Natl Med Assoc. 2018.
12. Santibanez M, Garrastazu R, Ruiz-Nunez M, Helguera JM, Arenal S, Bonnardeux C, et al. Predictors of Hospitalized Exacerbations and Mortality in Chronic Obstructive Pulmonary Disease. PLoS One. 2016;11(6):e0158727.
13. Rodriguez-Roisin R, Rabe KF, Vestbo J, Vogelmeier C, Agusti A, all p, et al. Global Initiative for Chronic Obstructive Lung Disease (GOLD) 20th Anniversary: a brief history of time. Eur Respir J. 2017;50(1).
14. Rabe KF, Watz H. Chronic obstructive pulmonary disease. Lancet. 2017;389(10082):1931-40.
15. Agusti A, Sin DD. Biomarkers in COPD. Clin Chest Med. 2014;35(1):131-41.
16. Faner R, Tal-Singer R, Riley JH, Celli B, Vestbo J, MacNee W, et al. Lessons from ECLIPSE: a review of COPD biomarkers. Thorax. 2014;69(7):666-72.
17. Goncalves I, Guimaraes MJ, van Zeller M, Menezes F, Moita J, Simao P, et al. Clinical and molecular markers in COPD. Pulmonology. 2018.
18. Agusti A, Bafadhel M, Beasley R, Bel EH, Faner R, Gibson PG, et al. Precision medicine in airway diseases: moving to clinical practice. Eur Respir J. 2017;50(4).

19. Agusti A, Celli B, Faner R. What does endotyping mean for treatment in chronic obstructive pulmonary disease? Lancet. 2017;390(10098):980-7.
20. Agusti A, Gea J, Faner R. Biomarkers, the control panel and personalized COPD medicine. Respirology. 2016;21(1):24-33.
21. Jones PW, Agusti AG. Outcomes and markers in the assessment of chronic obstructive pulmonary disease. Eur Respir J. 2006;27(4):822-32.
22. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary. Am J Respir Crit Care Med. 2017;195(5):557-82.
23. Hardy SE, Allore H, Studenski SA. Missing data: a special challenge in aging research. J Am Geriatr Soc. 2009;57(4):722-9.
24. Slatkin M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. Nat Rev Genet. 2008;9(6):477-85.
25. Wray NR. Allele frequencies and the r2 measure of linkage disequilibrium: impact on design and interpretation of association studies. Twin Res Hum Genet. 2005;8(2):87-94.
26. Lewontin RC. On measures of gametic disequilibrium. Genetics. 1988;120(3):849-52.
27. Chen Y, Lin CH, Sabatti C. Volume measures for linkage disequilibrium. BMC Genet. 2006;7:54.
28. Devlin B, Risch N. A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics. 1995;29(2):311-22.
29. Teare MD, Dunning AM, Durocher F, Rennart G, Easton DF. Sampling distribution of summary linkage disequilibrium measures. Ann Hum Genet. 2002;66(Pt 3):223-33.
30. Tenesa A, Wright AF, Knott SA, Carothers AD, Hayward C, Angius A, et al. Extent of linkage disequilibrium in a Sardinian sub-isolate: sampling and methodological considerations. Hum Mol Genet. 2004;13(1):25-33.
31. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet. 2017;101(1):5-22.
32. Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. Nat Rev Genet. 2014;15(5):335-46.
33. Wyss AB, Sofer T, Lee MK, Terzikhan N, Nguyen JN, Lahousse L, et al. Multiethnic meta-analysis identifies ancestry-specific and cross-ancestry loci for pulmonary function. Nat Commun. 2018;9(1):2976.
34. Terzikhan N, Sun F, Verhamme FM, Adams HHH, Loth D, Bracke KR, et al. Heritability and genome-wide association study of diffusing capacity of the lung. Eur Respir J. 2018.
35. Manchia M, Cullis J, Turecki G, Rouleau GA, Uher R, Alda M. The impact of phenotypic and genetic heterogeneity on results of genome wide association studies of complex diseases. PLoS One. 2013;8(10):e76295.
36. Github. Genetic correaltions ldscore 2017 [updated 22-04-2017. Available from: https://github.com/bulik/ldsc/wiki/FAQ.
37. Hackinger S, Zeggini E. Statistical methods to detect pleiotropy in human complex traits. Open Biol. 2017;7(11).
38. Ligthart S, Vaez A, Hsu YH, Inflammation Working Group of the CC, Pmi Wg XCP, LifeLines Cohort S, et al. Bivariate genome-wide association study identifies novel pleiotropic loci for lipids and inflammation. BMC Genomics. 2016;17:443.
39. Schaid DJ, Tong X, Larrabee B, Kennedy RB, Poland GA, Sinnwell JP. Statistical Methods for Testing Genetic Pleiotropy. Genetics. 2016;204(2):483-97.

Erasmus University Rotterdam

40. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nat Rev Genet. 2011;12(8):529-41.

41. Saffari A, Silver MJ, Zavattari P, Moi L, Columbano A, Meaburn EL, et al. Estimation of a significance threshold for epigenome-wide association studies. Genet Epidemiol. 2018;42(1):20-33.

42. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet. 2010;11(6):446-50.

43. Blanco-Gomez A, Castillo-Lluva S, Del Mar Saez-Freire M, Hontecillas-Prieto L, Mao JH, Castellanos-Martin A, et al. Missing heritability of complex diseases: Enlightenment by genetic variants from intermediate phenotypes. Bioessays. 2016;38(7):664-73.

44. Nedeljkovic I, Terzikhan N, Vonk JM, van der Plaat DA, Lahousse L, van Diemen CC, et al. A Genome-Wide Linkage Study for Chronic Obstructive Pulmonary Disease in a Dutch Genetic Isolate Identifies Novel Rare Candidate Variants. Front Genet. 2018;9:133.

45. Rosenkranz S, Gibbs JS, Wachter R, De Marco T, Vonk-Noordegraaf A, Vachiery JL. Left ventricular heart failure and pulmonary hypertension. Eur Heart J. 2016;37(12):942-54.

46. GOLD. Gobal strategy for the diagnosis, management and prevention of chronic obstructive pulmonary disease. 2018 report. 2018.

47. Bazan IS, Fares WH. Pulmonary hypertension: diagnostic and therapeutic challenges. Ther Clin Risk Manag. 2015;11:1221-33.

48. Wells JM, Washko GR, Han MK, Abbas N, Nath H, Mamary AJ, et al. Pulmonary arterial enlargement and acute exacerbations of COPD. N Engl J Med. 2012;367(10):913-21.

49. Zouk AN, Wells JM. In Rotterdam, size really does matter: implications of pulmonary artery enlargement on mortality. Eur Respir J. 2017;49(6).

50. Wells JM, Dransfield MT. Pathophysiology and clinical implications of pulmonary arterial enlargement in COPD. Int J Chron Obstruct Pulmon Dis. 2013;8:509-21.

51. Terzikhan N, Bos D, Lahousse L, Wolff L, Verhamme KMC, Leening MJG, et al. Pulmonary artery to aorta ratio and risk of all-cause mortality in the general population: the Rotterdam Study. Eur Respir J. 2017;49(6).

52. Townend D. Conclusion: harmonisation in genomic and health data sharing for research: an impossible dream? Hum Genet. 2018;137(8):657-64.

53. Knoppers BM, Joly Y. Introduction: the why and whither of genomic data sharing. Hum Genet. 2018;137(8):569-74.

54. Nyren O, Stenbeck M, Gronberg H. The European Parliament proposal for the new EU General Data Protection Regulation may severely restrict European epidemiological research. Eur J Epidemiol. 2014;29(4):227-30.

55. Chassang G. The impact of the EU general data protection regulation on scientific research. Ecancermedicalscience. 2017;11:709.

56. Morrison M, Bell J, George C, Harmon S, Munsie M, Kaye J. The European General Data Protection Regulation: challenges and considerations for iPSC researchers and biobanks. Regen Med. 2017;12(6):693-703.

57. Haug CJ. Turning the Tables - The New European General Data Protection Regulation. N Engl J Med. 2018;379(3):207-9.

58. Knoppers BM. Framework for responsible sharing of genomic and health-related data. Hugo J. 2014;8(1):3.

59. Rahimzadeh V, Dyke SO, Knoppers BM. An International Framework for Data Sharing: Moving Forward with the Global Alliance for Genomics and Health. Biopreserv Biobank. 2016;14(3):256-9.

60. Terry SF. The global alliance for genomics & health. Genet Test Mol Biomarkers. 2014;18(6):375-6.

61. Knoppers BM. International ethics harmonization and the global alliance for genomics and health. Genome Med. 2014;6(2):13.

62. Lawler M, Siu LL, Rehm HL, Chanock SJ, Alterovitz G, Burn J, et al. All the World's a Stage: Facilitating Discovery Science and Improved Cancer Care through the Global Alliance for Genomics and Health. Cancer Discov. 2015;5(11):1133-6.

63. consortium E. ENIGMA consortium: Enhancing neuroimaging genetics through meta-analysis 2018 [Available from: http://enigma.ini.usc.edu/.

64. consortium C. CHARGE consortium: cohorts for heart and aging research in genomic epidemiology 2018 [Available from: http://www.chargeconsortium.com/.

65. consortium G. GIANT: Genetic investigation of anthropometric triats 2018 [Available from: https://portals.broadinstitute.org/collaboration/giant/index.php/Main_Page.

66. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Int J Surg. 2014;12(12):1495-9.

67. Thelle DS. STROBE and STREGA: instruments for improving transparency and quality of reporting scientific results. Eur J Epidemiol. 2009;24(1):7-8.

68. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48(10):1279-83.

69. van de Vrugt HJ, Cornel MC, Wolthuis RMF. [CRISPR/Cas: technique to repair DNA errors: is a clinical breakthrough near?] CRISPR/Cas: techniek om DNA-fouten te repareren. Ned Tijdschr Geneeskd. 2018;162.

70. Center DM, Schwartz DA, Solway J, Gail D, Laposky AD, Lin QS, et al. Genomic medicine and lung diseases. Am J Respir Crit Care Med. 2012;186(3):280-5.

71. Okpechi SC, Ghonim MA, Lammi MR. Advances in Chronic Obstructive Pulmonary Disease Therapy: A Vascular-Targeted Approach. Clin Med Insights-Th. 2017;9.

72. Voelkel NF, Gomez-Arroyo J, Mizuno S. COPD/emphysema: The vascular story. Pulm Circ. 2011;1(3):320-6.