Approches en génomique et bio-informatique afin de comprendre les bases moléculaires de la maladie pulmonaire obstructive chronique

Thèse

Maxime Lamontagne

Doctorat en Médecine Expérimentale

Philosophiæ doctor (Ph.D.)

Québec, Canada

© Maxime Lamontagne, 2017

Approches en génomique et bio-informatique afin de comprendre les bases moléculaires de la maladie pulmonaire obstructive chronique

Thèse

Maxime Lamontagne

Sous la direction de :

Yohan Bossé, directeur de recherche

Résumé

La maladie pulmonaire obstructive chronique (MPOC) est une maladie complexe caractérisée par une obstruction non complètement réversible des voies aériennes. Actuellement, aucun remède n'existe pour soigner cette maladie qui deviendra la 3e cause de mortalité d'ici 2030. Bien que des progrès importants furent réalisés dans les dernières décennies, la pathogenèse de la maladie demeure largement mal comprise.

Cette thèse a pour objectif général l'étude des composantes génétiques de la MPOC, plus précisément 1) l'identification de gènes impliqués dans le développement de l'obstruction respiratoire, 2) l'identification d'eQTL pulmonaires dans le complexe majeur d'histocompatibilité et de leurs liens avec certains phénotypes pulmonaires, 3) l'identification de nouveaux gènes de susceptibilité au développement de la MPOC et 4) la caractérisation des variations génétiques du gène *SERPINA1* dans la population canadienne et l'évaluation du séquençage de l'ADN comme nouveau test diagnostique pour détecter le déficit en alpha-1 antitrypsine.

Dans une première étude, nous avons identifié des gènes (*CST3*, *CD22*) et des voies biologiques (métabolisme des xénobiotiques, apoptose, balance protéase-antiprotéase et balance oxydant-antioxydant) impliqués dans le développement de l'obstruction respiratoire. Pour ce faire, nous avons combiné les génotypes, les valeurs d'expression et les caractéristiques cliniques de 1,111 sujets. En combinant ces trois types de données, nous avons trouvé des variations génétiques influençant l'expression de gènes qui sont impliqués dans la perte de fonctions respiratoires. Ce projet a permis d'identifier des mécanismes précis impliqués dans le développement de l'obstruction respiratoire.

Par la suite, nous avons étudié une région clé du génome humain pour le système immunitaire, soit le complexe majeur d'histocompatibilité (MHC). Des études antérieures ont associé des variations génétiques présentes dans ce locus à plusieurs phénotypes pulmonaires (asthme, cancer du poumon, fibrose kystique, maladie pulmonaire interstitielle, MPOC). Les résultats obtenus nous ont permis d'identifier des gènes d'intérêts pour l'asthme (*AGPAT1*, *CDSN*), le cancer du poumon (*BTN3A2*, *ZFP57*), les fonctions respiratoires (*MICB*) et la maladie pulmonaire interstitielle (*AGPAT1*). Les

résultats de cette étude ont permis de mieux comprendre et d'interpréter les résultats obtenus précédemment.

Nous avons également participé à la plus grande étude pangénomique (GWAS) sur la MPOC. Cette étude, réalisée par l'*International COPD Genetics Consortium* (ICGC), a identifié 22 régions chromosomiques impliquées dans le développement de la MPOC. Pour ce projet, les génotypes de 63,192 sujets (47,936 cas et 15,256 témoins) provenant de 26 cohortes différentes ont été combinés dans une méta-analyse. Par la suite, les résultats ont été répliqués dans 9,498 cas et 9,748 témoins supplémentaires provenant de la *UK Biobank*. Parmi les 22 loci identifiés, 15 ont été précédemment associés avec les fonctions respiratoires et quatre n'avaient jamais été observés (*EEFSEC*, *DSP*, *MTCL1*, et *SFTPD*). Cette étude a permis de mieux comprendre l'architecture génétique de la maladie et de confirmer le lien existant entre les gènes associés aux fonctions respiratoires et la MPOC.

Finalement, nous avons caractérisé les variations génétiques du gène *SERPINA1* dans la population canadienne et testé le séquençage de l'ADN comme outil potentiel de diagnostic du déficit en alpha-1 antitrypsine. Pour ce faire, tous les exons codants du gène ont été séquencés dans 400 individus du projet pancanadien *CanCOLD* (*Canadian Cohort of Obstructive Lung Disease*). Nous avons identifié 19 variations génétiques dans le gène, dont 15 mutations faux sens et une nouvelle mutation. Une fois l'extraction d'ADN complété, le séquençage s'est effectué et a permis de déterminer de façon définitive la nature de l'anomalie génétique sous-jacente. Ce projet a démontré que cette technique permettrait d'accélérer le diagnostic et de diminuer les coûts pour ainsi faciliter la prise en charge et le traitement des patients souffrant d'un déficit en alpha-1 antitrypsine.

Les résultats obtenus au cours de ce doctorat permettent de mieux comprendre les gènes et les voies biologiques impliqués dans le développement de la MPOC.

Abstract

Chronic obstructive pulmonary disease (COPD) is a complex disease characterized by airflow obstruction that is not fully reversible. Currently, no treatment exists to reverse COPD, which is predicted to be the third leading cause of mortality in the world by the year 2030. Important discoveries were made in the last decade, but the pathophysiology of the disease remains largely unknown.

The aim of this thesis is to study the genetic component of COPD and more specifically 1) identify genes involved in the development of airflow obstruction, 2) identify lung eQTL in the major histocompatibility complex and find causal genes for lung function and respiratory diseases in this region, 3) find new susceptibility loci for COPD, and 4) evaluate the feasibility and effectiveness of DNA sequencing of the *SERPINA1* gene as a single test to diagnose alpha-1 antitrypsin deficiency (AATD) and test the frequencies of AATD alleles in a Canadian COPD population.

In the first study, we identified genes (*CST3* and *CD22*) and signalling pathways (xenobiotic metabolism, apoptosis, protease–antiprotease and oxidant–antioxidant balance) involved in the development of airflow obstruction. We combined lung gene expression, whole genotyping data and clinical information's from 1,111 subjects to identify potential causal genes and pathways. This study has identified underlying mechanisms implicated in the development of airflow obstruction.

In the second study, we studied a critical genomic region for the immune system, the major histocompatibility complex (MHC). Previous studies have associated single nucleotide polymorphisms (SNPs) located inside this locus with lung diseases and phenotypes (asthma, cystic fibrosis, idiopathic interstitial pneumonia, lung cancer and lung function). We have identified new susceptibility genes for lung cancer (*BTN3A2* and *ZFP57*), asthma (*AGPAT1* and *CDSN*), lung function (*MICB*) and idiopathic interstitial pneumonia (*AGPAT1*). Results from this study provide important biological insights about previously associated SNPs in the MHC.

We were also involved in the largest genome-wide association study (GWAS) on COPD. This GWAS was performed by the International COPD Genetics Consortium (ICGC) and identified 22 loci associated at genome-wide significance. Genotypes of 63,192 subjects (15,256 cases and 47,936 controls) from 26 studies were used in the meta-analysis. Results were further replicated in 9,498 cases and 9,748 controls from the UK Biobank. Among the 22 associated loci, 9 were previously associated with COPD, 15 with lung function and 4 (*EEFSEC*, *DSP*, *MTCL1* and *SFTPD*) were novel loci. Our findings highlight new loci associated with COPD and demonstrate the genetic overlap between lung function and COPD.

Finally, the frequencies of deficient *SERPINA1* alleles were evaluated in Canadian patients with COPD and DNA sequencing was evaluated as a single test strategy to detect AATD. DNA sequencing of the coding regions of *SERPINA1* was performed in 400 individuals from the CanCOLD study (Canadian Cohort of Obstructive Lung Disease). Nineteen genetic variants were identified, including 15 missense mutations and one new mutation. DNA sequencing of *SERPINA1* revealed the true genetic nature of AATD and was demonstrated has an effective, fast, and inexpensive single test strategy to detect AATD.

Studies presented in this thesis have identified genes and pathways involved in the development of COPD, which are new targets for future studies.

Table des matières

Résumé III
Abstract V
Table des matières VII
Liste des tableaux XI
Liste des figures XII
Liste des publications XIV
Liste des abréviations et sigles XX
Remerciements XXV
Avant-Propos XXVII
CHAPITRE 1 Introduction 1
1.1 : Maladie Pulmonaire Obstructive Chronique 2
1.1.1 Historique de la MPOC 3
1.1.2 Définitions et descriptions 6
1.1.2.1 Bronchite chronique 6
1.1.2.2 Emphysème 6
1.1.3 Symptômes et diagnostic de la MPOC 8
1.1.3.1 Diagnostic 8
1.1.3.2 Symptômes principaux 10
1.1.3.3 Exacerbations 11
1.1.3.4 Comorbidités 11
1.1.4 Prévalence et impact économique 15
1.1.5 Causes 17
1.1.5.1 Tabagisme 18
1.1.5.1.1 Immunité innée 18
1.1.5.1.2 Immunité adaptive 19
1.1.5.2 Déficit en alpha-1 antitrypsine 21
1.1.5.3 Autres causes 23
1.1.6 Traitements 24
1.1.6.1 Traitements non pharmacologiques 24
1.1.6.2 Traitements pharmacologiques 25
1.1.6.3 Traitement du déficit en alpha-1 antitrypsine 26
1.2 : Génétique de la maladie pulmonaire obstructive chronique 27
1.2.1 Introduction à la génétique 27
1.2.1.1 Acide désoxyribonucléique 27
1.2.1.2 Chromosome 28
1.2.1.3 Gène 30
1.2.1.4 ADN codant 31
1.2.1.5 Variation génétique 32
1.2.1.6 Technologies et approches en génomique 35
1.2.1.6.1 Puce à ADN (microarray) 35
1.2.1.6.2 Séquençage de l'ADN 35
1.2.1.6.3 Genome-wide association study (GWAS) 36
1.2.1.6.4 Expression Quantitative Trait Loci (eQTL) 38

1.2.2 Génétique de la MPOC 40
1.2.2.1 Déficit en alpha-1 antitrypsine 41
1.2.2.2 Études par gènes candidats 42
1.2.2.3 Genome-wide association study 43
1.2.2.3.1 4q22 – FAM13A 43
1.2.2.3.2 4q31 – HHIP 44
1.2.2.3.3 6p21.32 – AGER 44
1.2.2.3.4 15q25 – CHRNA3, CHRNA5, CHRNB4, IREB2 44
1.2.2.3.5 Fonction respiratoire 45
1.2.2.4 expression Quantitative Trait Loci 47
1.2.2.4.1 Consortium d'eQTL pulmonaire 47
1.2.2.4.2 Génétique de l'expression des gènes dans les expectorations de patients MPOC 47
1.2.2.4.3 Contrôle génétique de l'expression des gènes dans les loci de la MPOC48
CHAPITRE 2 Problématiques, objectifs et hypothèses des travaux 49
CHAPITRE 3 Genetic regulation of gene expression in the lung identifies CST3 and CD22 as potential causal genes for airflow obstruction 54
Résumé 57
Abstract 58
Key messages 59
Introduction 60
Methods 61
Results 63
Discussion 66
Acknowledgments 70
Contributorship 70
Competing interests 70
References 72
Tables 76
Figures 81
Online Data Supplement 83
Methods 83
Tables 88
Figures 90
CHAPITRE 4 Susceptibility genes for lung diseases in the major histocompatibility complex revealed by lung expression quantitative trait loci analysis 109
Résumé 112
Abstract 113
Introduction 114
Methods 115
Results 118
Discussion 120
Acknowledgements 123
Support 123
References 124
Tables 130
Figures 133

Supplementary data 136
Tables 136
Figures 137
CHAPITRE 5 Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis 141
Résumé 148
Abstract 149
Letter 150
Data Availability Statement 156
Author Contributions 156
Competing Financial Interest Statements : 157
References 159
Table 164
Figures 165
Supplement 169
Online Methods 169
References 177
Supplementary Materials 180
CHAPITRE 6 Comprehensive genetic testing for alpha-1 antitrypsin deficiency in COPD patients 181
Funding information 183
Abbreviations 183
Résumé 184
Abstract 185
Introduction 186
Materials and Methods 187
Results 188
Author contributions 193
Financial/nonfinancial disclosures 193
Role of the sponsors 193
References 194
Tables 196
Figures 200
CHAPITRE 7 Discussion, conclusion et perspectives 202
7.1 Obstruction respiratoire 203
7.1.1 Modèle de causalité 204
7.1.2 Voies biologiques 207
7.1.3 Conclusion partielle 1 207
7.2 Complexe majeur d'histocompatibilité 208
7.2.1 Génome 208
7.2.2 Gènes causaux dans la région du MHC 209
7.2.3 Conclusion partielle 2 213
7.3 Régions de susceptibilité 213
7.3.1 Méta-analyse 213
7.3.2 Corrélation génétique 215
7.3.3 Conclusion partielle 3 215
7.4 Déficit en alpha-1 antitrypsine 216

7.4.1 Séquençage 216
7.4.2 Temps et coûts 217
7.4.3 Conclusion partielle 4 217
7.5 Conclusion et perspectives 217
Références 219

Liste des tableaux

Tableau 1-1 : Symptômes cliniques à considérer pour le diagnostic de la MPOC. 9
Tableau 1-2 : Classification des stades d'obstruction respiratoire. 9
Tableau 1-3 : Caractéristiques des comorbidités associées à la MPOC. 13
Tableau 1-4 : Code de l'ADN. 32
Tableau 1-5 : GWAS effectués sur la MPOC. 43
Table 3-1 : Clinical characteristics of patients. 76
Table 3-2 : Models of biological relevance identified by manual curation. 77
Table 3-3 : Canonical pathways enriched for causality genes using the Ingenuity Pathway Analysis system. 78
Table 3-4 : Canonical pathways enriched for causality genes using Partek. 79
Table 3-5 : Replication of causality genes in the bronchial airway epithelial and the regional lung tissue datasets. 80
Table 3-S1 : List of causal models with a confidence score >0.8. 88
Table 3-S2 : Models of biological relevance identified by manual curation 88
Table 4-1 : Clinical characteristics of patients that passed gene expression and genotyping quality control filters 130
Table 4-2 : Significant lung eQTLs detected in the xMHC for GWAS associated SNPs with lung diseases 131
Table 4-S1 : GWAS SNPs with P-values < 1 x 10-5 136
Table 4-S2 : Number of local-eQTL-SNPs per window of 20 Mb. 136
Table 5-1 : Overall study results showing 22 loci with genome-wide significant P values in overall meta-analysis following UK BiLEVE Stage 2 analysis. 164
Table 6-1 : Clinical characteristics of COPD patients. 196
Table 6-2 : Primers to sequence the SERPINA1 gene. 197
Table 6-3 : Time and cost to sequence the coding regions of SERPINA1 in batch of 24 patients. 198
Table 6-4 : Nomenclature of the 19 genetic variants identified and allele frequencies in this Canadian COPD population. 199

Liste des figures

Figure 1-1 : Représentation graphique d'un poumon emphysémateux datant de 1803. 4
Figure 1-2 : Différences entre un poumon sain et un poumon de patient MPOC. 7
Figure 1-3 : Déclin du VEMS chez les fumeurs susceptibles. 10
Figure 1-4 : Comorbidités associées à la MPOC. 12
Figure 1-5 : Prévalence de la MPOC selon le groupe d'âge et le sexe aux États-Unis entre 2007 et 2009. 17
Figure 1-6 : Structure de l'ADN. 28
Figure 1-7 : Structure d'un chromosome autosomal. 29
Figure 1-8 : Structure d'un gène. 30
Figure 1-9 : Transcription et traduction. 31
Figure 1-10 : Types de polymorphisme nucléotidique simple (SNP) codant. 33
Figure 1-11 : Exemples d'insertions et de délétions. 34
Figure 1-12 : Exemple de criblage génomique par association (GWAS). 37
Figure 1-13 : Exemple d'expression Quantitative Trait Loci. 38
Figure 1-14 : eQTL local et distant. 40
Figure 1-15 : Graphique de Manhattan de Wain et al., Nat Gen, 2017. 46
Figure 3-1 : Causality models showing the relationships between the expression of a gene, a phenotype, and a SNP. 81
Figure 3-2 : Direction of effects for causality model with rs6048956, mRNA expression of CST3, and FEV1 % predicted. 82
Figure 3-S1 : Direction of effects for causality model with rs6515375, mRNA expression of CST3, and FEV1 % predicted. Data are presented as described in Figure 3-2. 90
Figure 3-S2 : Direction of effects for causality model with rs769178, mRNA expression of NCR3, and FEV1 % predicted. Data are presented as described in Figure 3-2. 91
Figure 3-S3 : Direction of effects for causality model with rs2270859, mRNA expression of CSTA, and FEV1 % predicted. Data are presented as described in Figure 3-2. 92
Figure 3-S4 : Direction of effects for causality model with rs4550905, mRNA expression of PPARGC1A, and FEV1 % predicted. Data are presented as described in Figure 3-2. 93
Figure 3-S5 : Direction of effects for causality model with rs3803761, mRNA expression of FLCN, and FEV1 % predicted. Data are presented as described in Figure 3-2. 94
Figure 3-S6 : Direction of effects for causality model with rs1543438, mRNA expression of BCL2L1, and FEV1 % predicted. Data are presented as described in Figure 3-2. 95
Figure 3-S7 : Direction of effects for causality model with rs9880397, mRNA expression of CADM2, and FEV1 % predicted. Data are presented as described in Figure 3-2. 96
Figure 3-S8 : Direction of effects for causality model with rs2466183, mRNA expression of TNFRSF10B, and FEV1 % predicted. Data are presented as described in Figure 3-2. 97
Figure 3-S9 : Direction of effects for causality model with rs17754977, mRNA expression of GSTO2, and FEV1/FVC. Data are presented as described in Figure 3-2. 98

Figure 3-S10 : Direction of effects for causality model with rs9987135, mRNA expression of DEPDC6, and FEV1/FVC. Data are presented as described in Figure 3-2. 99
Figure 3-S11 : Direction of effects for causality model with rs10411704, mRNA expression of CD22, and FEV1/FVC. Data are presented as described in Figure 3-2. 100
Figure 3-S12 : Direction of effects for causality model with rs12179536, mRNA expression of MUC22, and FEV1/FVC. Data are presented as described in Figure 3-2. 101
Figure 3-S13 : Direction of effects for causality model with rs2287765, mRNA expression of SPINK5, and FEV1/FVC. Data are presented as described in Figure 3-2. 102
Figure 3-S14 : Causality genes in the Aryl Hydrocarbon Receptor Signaling pathway. Causality genes identified in this study are in red. 103
Figure 15 : Causality genes in the Xenobiotic Metabolism Signaling pathway. Causality genes identified in this study are in red. 103
Figure 3-16 : The glutathione metabolism pathway connecting the cyanoamino acid and the taurine/hypotaurine metabolism pathways. 104
Figure 3-S17 : Gene Ontology (GO) project functional categories enriched for causality genes. 105
Figure 4-1: Number of local-eQTL-SNPs per sliding-window of 20 Mb. 133
Figure 4-2 : Lung eQTL in the xMHC for GWAS-nominated SNPs of lung diseases and traits. 135
Figure 4-S1 : Linkage disequilibrium plot of selected GWAS SNPs. 137
Figure 4-S2 : The effect of distance between the eQTL-SNP and the transcription start site (TSS) of the eQTL-gene in the discovery cohort (Laval). 138
Figure 4-S3 : Boxplots of ZFP57 gene expression levels in the lung according to genotype groups of rs375984. 139
Figure 4-S4 : Circos plot of lung eQTLs located in the xMHC. 140
Figure 5-1 : Study design showing the cohorts used in each stage of the analysis. 165
Figure 5-2 : Manhattan plot. 166
Figure 5-3 : Regional association for novel loci. 167
Figure 5-4 : Genetic correlation (using LD score regression) between COPD and other traits. 168
Figure 6-1 : The exon-intron structure of the SERPINA1 gene and the localization of the identified genetic variants. 200
Figure 6-2 : Number of patients identified in any pairwise combinations of normal and deficient alleles annotated based on the legacy nomenclature. 201

Liste des publications

1. **Maxime Lamontagne**, Wim Timens, Ke Hao, Yohan Bossé, Michel Laviolette, Katrina Steiling, Joshua D Campbell, Christian Couture, Massimo Conti, Karen Sherwood, James C Hogg, Corry-Anke Brandsma, Maarten van den Berge, Andrew Sandford, Stephen Lam, Marc E Lenburg, Avrum Spira, Peter D Paré, David Nickle, Don D Sin, Dirkje S Postma. Genetic regulation of gene expression in the lung identifies CST3 and CD22 as potential causal genes for airflow obstruction. Thorax. 2014 Nov;69(11):997-1004. PMID 25182044.

2. **Maxime Lamontagne**, Philippe Joubert, Wim Timens, Dirkje S. Postma, Ke Hao, David Nickle, Don D. Sin, Peter D. Pare, Michel Laviolette and Yohan Bossé. Susceptibility genes for lung diseases in the major histocompatibility complex revealed by lung expression quantitative trait loci analysis. Eur Respir J. 2016 Aug;48(2):573-6. PMID 27174877.

3. Brian D. Hobbs, Kim de Jong, **Maxime Lamontagne**, Yohan Bossé, Nick Shrine, María Soler Artigas, Victoria E. Jackson, Louise V. Wain, Ian P. Hall, Annah B. Wyss, Stephanie J. London, Kari E. North, Nora Franceschini, David P. Strachan, Terri H. Beaty, John E. Hokanson, James D. Crapo, Peter J. Castaldi, Robert P. Chase, Traci M. Bartz, Susan R. Heckbert, Bruce M. Psaty, Sina A. Gharib, Pieter Zanen, Jan W. Lammers, Matthijs Oudkerk, H. J. Groen, Nick Locantore, Ruth Tal-Singer, Stephen I. Rennard, Wim Timens, Peter D. Paré, Jeanne C. Latourelle, Josée Dupuis, George T. O'Connor, Jemma B. Wilk, Woo Jin Kim, Mi Kyeong Lee, Yeon-Mok Oh, Judith M. Vonk, Harry J. de Koning, Shuguang Leng, Steven A. Belinsky, Yohannes Tesfaigzi, Ani Manichaikul, Xin-Qun Wang, Stephen S. Rich, R Graham Barr, David Sparrow, Augusto A. Litonjua, Per Bakke, Amund Gulsvik, Lies Lahousse, Guy G. Brusselle, Bruno H. Stricker, André G. Uitterlinden, Elizabeth J. Ampleford, Eugene R. Bleecker, Prescott G. Woodruff, Deborah A. Meyers, Dandi Qiao, David A. Lomas, Jae-Joon Yim, Deog Kyeom Kim, Iwona Hawrylkiewicz, Pawel Sliwinski, Megan Hardin, Tasha E. Fingerlin, David A. Schwartz, Dirkje S. Postma, William MacNee, Martin D. Tobin, Edwin K.

Silverman, H. Marike Boezen, Michael H. Cho, COPDGene Investigators, ECLIPSE Investigators, LifeLines Investigators, SPIROMICS Research Group, International COPD Genetics Network Investigators, UK BiLEVE Investigators, International COPD Genetics Consortium.. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. Nat Genet. 2017 Mar;49(3):426-432. PMID 28166215.

4. **Maxime Lamontagne**, Nathalie Gaudreault, François Maltais, Wan Tan, Jean Bourbeau, Yohan Bossé, for the Canadian Respiratory Research Network, for the CanCOLD study group. Comprehensive genetic testing for alpha-1 antitrypsin deficiency in COPD patients. In preparation.

Autres publications comme co-auteur réalisées durant ma thèse : :

1. Ma'en Obeidat, Suzanne Miller, Kelly Probert, Charlotte K. Billington, Amanda P. Henry, Emily Hodge, Carl P. Nelson, Ceri E. Stewart, Caroline Swan, Louise V. Wain, María Soler Artigas, Erik Melén, Kevin Ushey, Ke Hao, **Maxime Lamontagne**, Yohan Bossé, Dirkje S. Postma, Martin D. Tobin, Ian Sayers, Ian P. Hall. GSTCD and INTS12 regulation and expression in the human lung. PLoS One. 2013 Sep 18;8(9):e74630. PMID 24058608.

2. Jean-Christophe Bérubé, **Maxime Lamontagne**, Christian Couture, David Nickle, Wim Timens, Dirkje S. Postma, Don D. Sin, Peter D. Paré, Michel Laviolette, Yohan Bossé. Genome-wide genetic ancestry measurements to predict lung function in European populations. Eur Respir J. 2013 Oct;42(4):1144-7. PMID 24081765

3. Mathieu C. Morissette, **Maxime Lamontagne**, Jean-Christophe Bérubé, Gordon Gaschler, Andrew Williams, Carole Yauk, Christian Couture, Michel Laviolette, James C. Hogg, Wim Timens, Sabina Halappanavar, Martin R. Stampfli, Yohan Bossé. Impact of Cigarette Smoke on the Human and Mouse Lungs: A Gene-Expression Comparison Study. PLoS One. 2014 Mar 24;9(3):e92498. PMID 24663285.

4. Justin Dang Uy Nguyen, **Maxime Lamontagne**, Christian Couture, Massimo Conti, Peter D. Paré, Don D. Sin, James C. Hogg, David Nickle, Dirkje S. Postma, Wim Timens, Michel Laviolette and Yohan Bossé. Susceptibility loci for lung cancer are associated with mRNA levels of nearby genes in the lung. Carcinogenesis. 2014 Dec;35(12):2653-9. PMID 25187487.

5. Frédéric Guénard, **Maxime Lamontagne**, Yohan Bossé, Yves Deshaies, Katherine Cianflone, John G. Kral, Picard Marceau, Marie-Claude Vohl. Influences of Gestational Obesity on Associations between Genotypes and Gene Expression Levels in Offspring following Maternal Gastrointestinal Bypass Surgery for Obesity. PLoS One. 2015 Jan 20;10(1):e0117011. PMID 25603303.

6. Aabida Saferali, Ma'en Obeidat, Jean-Christophe Bérubé, **Maxime Lamontagne**, Yohan Bossé, Michel Laviolette, Ke Hao, David C. Nickle, Wim Timens, Don D. Sin, Dirkje S. Postma, Lisa J. Strug, Paul J. Gallins, Peter D. Paré, Colin D. Bingle and Andrew J. Sandford. Polymorphisms Associated with Expression of BPIFA1/BPIFB1 and Lung Disease Severity in Cystic Fibrosis. Am J Respir Cell Mol Biol. 2015 Nov;53(5):607-14. PMID 25574903.

7. Sandra Guauque-Olarte, David Messika-Zeitoun, Arnaud Droit, **Maxime Lamontagne**, Joël Tremblay-Marchand, Emilie Lavoie-Charland, Nathalie Gaudreault, Benoit J. Arsenault, Marie-Pierre Dubé, Jean-Claude Tardif, Simon C. Body, Jonathan G. Seidman, Catherine Boileau, Patrick Mathieu, Philippe Pibarot, Yohan Bossé. Calcium Signaling Pathway Genes RUNX2 and CACNA1C Are Associated with Calcific Aortic Valve Disease. Circ Cardiovasc Genet. 2015 Dec;8(6):812-22. PMID 26553695.

8. Natasha Dargis, **Maxime Lamontagne**, Nathalie Gaudreault, Laura Sbarra, Cyndi Henry, Philippe Pibarot, DVM, Patrick Mathieu and Yohan Bossé. Identification of Gender-Specific Genetic Variants in Patients with Bicuspid Aortic Valve. Am J Cardiol. 2016 Feb 1;117(3):420-6. PMID 26708639.

9. Hans Matsson, Cilla Söderhäll, Elisabet Einarsdottir, **Maxime Lamontagne**, Sanna Gudmundsson, Helena Backman, Anne Lindberg, Eva Rönmark, Juha Kere, Don Sin, Dirkje S. Postma, Yohan Bossé, Bo Lundbäck, Joakim Klar. Targeted high-throughput sequencing of candidate genes for chronic obstructive pulmonary disease. BMC Pulm Med. 2016 Nov 11;16(1):146. PMID 27835950.

10. James D. McKay, Rayjean J. Hung, Younghun Han, Xuchen Zong, Robert Carreras-Torres, David C. Christiani, Neil Caporaso, Mattias Johansson, Xiangjun Xiao, Yafang Li, Jinyoung Byun, Alison Dunning, Karen Poole, David C. Qian, Xuemei Ji, Geoffrey Liu, Maria Timofeeva, Stig E. Bojesen, Xifeng Wu, Loic Le Marchand, Demetrios Albanes, Heike Bickeböller, Melinda C. Aldrich, William S. Bush, Adonina Tardon, Gad Rennert, M. Dawn Teare, John K. Field, Lambertus A. Kiemeney, Philip Lazarus, Aage Haugen, Stephen Lam, Matthew B. Schabath, Angeline S. Andrew, Hongbing Shen, Yun-Chul Hong, Jian-Min Yuan, Pier Alberto Bertazzi, Angela C. Pesator, Yuanqing Ye, Nancy Diao, Li Su, Ruyang Zhang, Yonathan Brhane, Natasha Leighl, Jakob S. Johansen, Anders Mellemgaard, Dr. Walid Saliba, Christopher Haiman, Lynne Wilkens, Ana Fernandez-Somoano, Guillermo Fernandez-Tardon, Henricus F.M. van der Heijden, Jin Hee Kim, Juncheng Dai, Zhibin Hu, Michael PA Davies, Michael W. Marcus, Hans Brunnström, Jonas Manjer, Olle Melander, David C. Muller, Kim Overvad, Antonia Trichopoulou, Rosario Tumino, Jennifer Doherty, Matt Barnett, Chu Chen, Gary Goodman, Angela Cox, Fiona Taylor, Penella Woll, Irene Brüske, Judith Manz, Thomas Muley, Angela Risch, Albert Rosenberger, Kjell Grankvist, Mikael Johansson, Frances A. Shepherd, Ming-Sound Tsao, Susanne M. Arnold, Eric B. Haura, Ciprian Bolca, Ivana Holcatova, Vladimir Janout, Milica Kontic, Jolanta Lissowska, Anush Mukeria, Simona Ognjanovic, Tadeusz M. Orlowski, Ghislaine Scelo, Beata Swiatkowska, David Zaridze, Per Bakke, Vidar Skaug, Shanbeh Zienolddiny, Eric J. Duell, Lesley M. Butler, Woon-Puay Koh, Yu-Tang Gao, Richard Houlston, John McLaughlin, Victoria Stevens, Philippe Joubert, **Maxime Lamontagne**, David C. Nickle, Ma'en Obeidat, Wim Timens, Bin Zhu, Lei Song, Linda Kachuri, María Soler Artigas, Martin D. Tobin, Louise V. Wain, SpiroMeta Consortium, Thorunn Rafnar, Thorgeir E. Thorgeirsson, Gunnar W. Reginsson,

Kari Stefansson, Dana B. Hancock, Laura J. Bierut, Margaret R. Spitz, Nathan C Gaddis, Sharon M. Lutz, Fangyi Gu, Eric O. Johnson, Ahsan Kamal, Claudio Pikielny, Dakai Zhu, Sara Lindströem, Xia Jiang, Rachel F. Tyndale, Georgia Chenevix-Trench, Jonathan Beesley, Yohan Bossé, Stephen Chanock, Paul Brennan, Maria Teresa Landi, Christopher I. Amos. Large scale genetic analysis identifies novel loci and histological variability in susceptibility to lung cancer. Nat Genet (Publication in principle).

11. Xuemei Ji, Yohan Bossé, Maria Landi, Jiang Gui, Xiangjun Xiao, David Qian, Philippe Joubert, Maxime Lamontagne, Yafang Li, Ivan Gorlov, Mariella Biasi, Younghun Han, Olga Gorlova, Rayjean Hung, Xifeng Wu, James McKay, Xuchen Zong, Robert Carreras-Torres, David Christiani, Neil Caporaso, Mattias Johansson, Geoffrey Liu, Stig Bojesen, Loic Le Marchand, Demetrius Albanes, Heike Bickeböller, Melinda Aldrich, William Bush, Adonina Tardon, Gad Rennert, Chu Chen, M. Dawn Teare, John Field, Lambertus Kiemeney, Philip Lazarus, Aage Haugen, Stephen Lam Lam, Matthew Schabath, Angeline Andrew, Hongbing Shen, Yun-Chul Hong, Jian-Min Yuan, Pietro Bertazzi, Angela Pesatori, Yuanqing Ye, Nancy Diao, Li Su, Ruyang Zhang, Yonathan Brhane, Natasha Leighl, Jakob Johansen, Anders Mellemgaard, Walid Saliba, Christopher Haiman, Lynn Wilkens, Ana Fernandez-Somoano, Guillermo Fernandez-Tardon, Erik Heijden, Jin Hee Kim, Juncheng Dai, Zhibin Hu, Michael Davies, Michael Marcus, Hans Brunnstrom, Jonas Manjer, Olle Melander, David Muller, Kim Overvad, Antonia Trichopoulou, Rosario Tumino, Jennifer Doherty, Gary Goodman, Angela Cox, Fiona Taylor, Penella Woll, Irene Bruske, Judith Manz, Thomas Muley, Angela Risch, Albert Rosenberger, Kjell Grankvist, Mikael Johansson, Frances Shepherd, Ming-Sound Tsao, Susanne Arnold, Eric Haura, Ciprian Bolca, Ivana Holcatova, Vladimir Janout, Milica Kontic, Jolanta Lissowska, Anush Mukeria, Simona Ognjanovic, Tadeusz Orlowski, Ghislaine Scelo, Beata Swiatkowska, David Zaridze, Per Bakke, Vidar Skaug, Shanbeh Zienolddiny, Eric Duell, Lesley Butler, Woon-Puay Koh, Yutang Gao, Richard Houlston, John McLaughlin, Victoria Stevens, David Nickle, Ma'en Obeidat, Wim Timens, Bin Zhu, Lei Song, María Artigas, Martin Tobin, Louise Wain, Fangyi Gu, Jinyoung Byun, Ahsan Kamal,

Dakai Zhu, Rachel Tyndale, Georgia Chenevix-Trench, Jonathan Beesley, Stephen Chanock, Paul Brennan. Identification of susceptibility pathways for chromosome 15q25.1 modifying lung cancer risk using genome-wide association analysis. Nat Com (submit).

Liste des abréviations et sigles

%	pourcentage
AA	African Americans
AAT	alpha-1 antitrypsine
AATD	alpha-1 antitrypsin deficiency
ACLA	anticholinergiques à longue durée d'action
ADAM19	ADAM metallopeptidase domain 19
ADGRG6	adhesion G protein-coupled receptor G6
ADN	acide désoxyribonucléique
AGER	advanced glycosylation end-product specific receptor
AGPAT1	1-acylglycerol-3-phosphate O-acyltransferase 1
Ala	alanine
ALDH8A1	aldehyde dehydrogenase 8 family member A1
Arg	arginine
ARIC	Atherosclerosis Risk Study in Communities
ARMC2	armadillo repeat containing 2
ARNm	acide ribonucléique messager
ARNT	aryl hydrocarbon receptor nuclear translocator
Asn	asparagine
Asp	acide aspartique
B58	British 1958 Birth Cohort
BALA	β2-adrénergiques à longue durée d'action
BCL2L1	BCL2 like 1
BDCA	bronchodilatateurs à courte durée d'action
BPCO	broncho-pneumopathie chronique obstructive
BTN3A2	butyrophilin subfamily 3 member A2
CADM2	cell adhesion molecule 2
CAE	condensats de l'air expiré
CanCOLD	the Canadian Cohort of Obstructive Lung Disease
CD22	CD22 molecule
CDSN	corneodesmosin
CEU	Northern Europeans from Utah
CFDP1	craniofacial development protein 1
CFTR	cystic fibrosis transmembrane conductance regulator
CHRNA3	cholinergic receptor nicotinic alpha 3 subunit
CHRNA5	cholinergic receptor nicotinic alpha 5 subunit
CHRNB4	cholinergic receptor nicotinic beta 4 subunit
CHS	Cardiovascular Health Study
CNV	copy number variation
COPACETIC	COPD Pathology : Addressing Critical Gaps, Early Treatment and Diagnosis and Innovative Concepts
COPD	chronic obstructive respiratory disease
CSI	corticostéroïdes en inhalation
CST3	cystatin C
CSTA	cystatin A

CVF	capacité vitale forcée
CYP2A6	cytochrome P450 family 2 subfamily A member 6
Cys	cystéine
DAAT	déficit en apha-1 antitrypsine
dbSNP	the single nucleotide polymorphism database
DEPDC6	DEP domain containing 6
dL	décilitre
DSP	desmoplakin
EA	European Americans
EBI	European Bioinformatics Institute
ECLIPSE	Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-Points
EEFSEC	eukaryotic elongation factor, selenocysteine-tRNA specific
EOCOPD	Boston Early-Onset COPD Study
eQTL	expression Quantitative Trait Loci
FAM13A	family with sequence similarity 13 member A
FEV1	forced expiratory volume in 1 second
FGF7	fibroblast growth factor 7
FHS	Framingham Heart Study
FLCN	folliculin
FLG	filaggrine
FRQS	Fonds de recherche du Québec – santé
FVC	forced vital capacity
g	gramme
GGT	gamma-glutamyl transferase
Gln	glutamine
Glu	acide glutamique
Gly	glycine
GOLD	global initiative for chronic obstructive lung disease
GPR126	Adhesion G protein-coupled receptor G6
GSTCD	glutathione S-transferase C-terminal domain containing
GSTM1	glutathione S-transferase mu 1
GSTO2	glutathione S-transferase omega 2
GWAS	genome-wide association study
HCG4B	HLA complex group 4B
hg19	human genome 19
HHIP	hedgehog interacting protein [Homo sapiens
His	histidine
HIST1H2AA	histone cluster 1 H2A family member a
HLA	human leukocyte antigen
HLA-A	major histocompatibility complex, class I, A
HLA-DPB1	major histocompatibility complex, class II, DP beta 1
HLA-DPB2	major histocompatibility complex, class II, DP beta 2
HLA-DQB1	major histocompatibility complex, class II, DQ beta 1
HLA-DQB2	major histocompatibility complex, class II, DQ beta 2
HLA-F	major histocompatibility complex, class I, F
HLA-J	major histocompatibility complex, class I, J

HTR4	5-hydroxytryptamine receptor 4
ICGC	international COPD genetics consortium
ICGN	International COPD Genetics Network
IFITM4P	interferon induced transmembrane protein 4 pseudogene
IFNA4	interferon alpha 4
IFNA5	interferon alpha 5
IFNA10	interferon alpha 10
IFNA13	interferon alpha 13
IFNA17	interferon alpha 17
IFNA21	interferon alpha 21
IFNγ	interféron gamma
IGF2	insulin like growth factor 2
IL6	interleukin 6
IL27	interleukin 27
Ile	isoleucine
Indel	insertion et délétion
IPA	ingenuity pathways analysis
IREB2	iron responsive element binding protein 2
IRF7	interferon regulatory factor 7
J.-C.	Jésus Christ
KARE	Korean Association Resource Project
Kb	kilo base
KIFC1	kinesin family member C1
L	Littre
LD	linkage disequilibrium
Leu	leucine
LPA	acide lysophosphatidique
Lys	lysine
MAC	minor allele count
Mb	méga base
MESA	Multi-Ethnic Study of Atherosclerosis
Met	methionine
mg	milligramme
MHC	major histocompatibility complex
MICB	MHC class I polypeptide-related sequence B
MPOC	maladie pulmonaire obstructive chronique
mRNA	messenger ribonucleic acid
MTCL1	microtubule crosslinking factor 1
mTORC1	mammalian target of rapamycin complex 1
mTORC2	mammalian target of rapamycin complex 2
MUC22	mucin 22
NCR3	natural cytotoxicity triggering receptor 3
NETT-NAS	National Emphysema Treatment Trial/Normative Aging Study
NHGRI	National Human Genome Research Institute
NHW	Non-Hispanic whites
NK	Natural killer
NLRX1	NLR family member X1

NMI	N-myc and STAT interactor
NOS1AP	nitric oxide synthase 1 adaptor protein
NPNT	nephronectin
OMS	organisation mondiale de la santé
P	valeur p
PA	acide phosphatidique
Phe	phénylalanine
Pi	protease inhibitor
PID1	phosphotyrosine interaction domain containing 1
PPARGC1A	PPARG coactivator 1 alpha
PRKCE	protein kinase C epsilon
Pro	proline
PSMA4	proteasome subunit alpha 4
PSORS1C1	psoriasis susceptibility 1 candidate 1
RARB	retinoic acid receptor beta
RIN3	Ras and Rab interactor 3
RNA	ribonucleic acid
ROS	reactive oxygen species
RS	Rotterdam Study
Ser	sérine
SERPINA1	serpin peptidase inhibitor clade A member 1
SFTPD	surfactant protein D
SNP	single-nucleotide polymorphism
SOGA2	SOGA family member 2
SPINK5	serine peptidase inhibitor, Kazal type 5
SPIROMICS	Subpopulations and Intermediate Outcome Measures in COPD Study
TAP2	transporter 2, ATP binding cassette subfamily B member
TCGS	Transcontinental COPD Genetics Study
TET2	tet methylcytosine dioxygenase 2
TGFB1	transforming growth factor beta 1
TGFB2	transforming growth factor beta 2
Thr	threonine
THSD4	thrombospondin type 1 domain containing 4
TNF-α	tumor necrosis factor alpha
TNFRSF10B	TNF receptor superfamily member 10b
TRAIL	TNF-related apoptosis-inducing ligand
TRAILR2	TRAIL receptor 2
TRIP11	thyroid hormone receptor interactor 11
Trp	tryptophane
TUFM	Tu translation elongation factor, mitochondrial
Tyr	tyrosine
UBC	University of British Columbia
UK BiLEVE	UK Biobank Lung Exome Variant Evaluation
Val	valine
VEMS	volume expiratoire maximal seconde
xMHC	extended major histocompatibility complex
ZFP57	ZFP57 zinc finger protein

ZSCAN31	zinc finger and SCAN domain containing 31

Remerciements

D'abord et avant tout, je veux remercier mon superviseur Yohan Bossé pour m'avoir donné la chance de faire partie de son équipe de recherche et de m'avoir encouragé (et supporté) tout au long de ces années.

Merci à Nathalie Gaudreault pour son support, ses réponses, son temps, ses desserts, etc. Merci pour tout.

Merci à Cyndi Henry et Laura Sbarra pour votre contribution à mon poids et autres niaiseries quotidiennes. Avec vous, les journées passaient aussi vite que vos desserts.

Benny, merci pour le cholestérol, le gaming et les appels téléphoniques matinaux. « Va ... toé pis toute ta famille ».

Merci à Alisson Clemenceau et Jean-Christophe Bérubé d'être présent, cela nous permet de partager notre souffrance collective.

Merci Emilie Bernatchez pour ta présence à la fin de ce long parcours.

Merci également à toutes les autres personnes ayant évoluées dans le laboratoire pendant ces nombreuses années : Natasha « Banana » Dargis, Sandra Guauque-Olarte, Émilie Lavoie-Charland, Gabrielle Prévost, Joël Tremblay-Marchand, Josée Vézina.

Je veux également remercier les organismes subventionnaires qui m'ont supporté tout au long de mes études supérieures : l'Université Laval, le Centre de recherche de l'Institut universitaire de cardiologie et de pneumologie de Québec (CRIUCPQ), les Fonds de recherche Québec – Santé (FRQS), l'Association des chercheuses et chercheurs étudiant à la Faculté de médecine de l'Université Laval (ACCEM). Grâce à leur support, j'ai pu évoluer dans un domaine que j'adorais avec un stress en moins.

Merci également à mes nombreux collaborateurs locaux (Christian Couture, Philippe Joubert, Michel Laviolette, François Maltais) et internationaux (Michael H. Cho, Ke Hao, Brian D. Hobbs, David Nickle, Peter D Paré, Dirkje S Postma, Don D Sin, Wim Timens) pour leurs projets stimulants et leurs commentaires constructifs.

Finalement, je veux remercier ma famille pour son support constant même s’ils ne comprenaient pas toujours ce que je faisais.

Avant-Propos

Cette thèse comprend 4 articles ayant comme sujet principal la génétique de la maladie pulmonaire obstructive chronique (MPOC). Le premier chapitre est une introduction générale de plusieurs concepts abordés dans la thèse (**Chapitre 1).** Ce chapitre est divisé en deux sections principales, la MPOC (historique, description, symptômes, prévalence, causes, traitements) et la génétique de la MPOC. Le **Chapitre 2** décrit les objectifs et les hypothèses de recherche du projet. Les 4 articles scientifiques suivront par la suite (**Chapitres 3 à 6**). À noter que les références de ces articles sont exclues des références finales de la thèse.

Le premier article de cette thèse s'intitule *Genetic regulation of gene expression in the lung identifies CST3 and CD22 as potential causal genes for airflow obstruction* (**Chapitre 3**). L'objectif était d'identifier des gènes et des voies biologiques régulés par les variations génétiques précédemment associées à la MPOC et aux fonctions respiratoires. Cette étude a démontré que les gènes *CST3* et *CD22* ainsi que les voies biologiques de l'apoptose, du métabolisme des xénobiotiques, et les balances protéase-antiprotéase et oxydant-antioxydant sont impliqués dans le développement de l'obstruction respiratoire. Cette étude a été réalisée dans le cadre du consortium d'eQTL pulmonaire. Mon rôle dans cette étude fut la réalisation d'analyses bio-informatiques, l'interprétation et la rédaction des résultats obtenus. Je suis l'un des trois auteurs principaux avec les Dr Wim Timens (Université de Groningen au Pays-Bas) et Ke Hao (Mount Sinai, New York, États-Unis). L'article a été soumis dans la revue *Thorax* le 28 avril 2014 et a été accepté pour publication le 14 août 2014.

Le deuxième article s'intitule *Susceptibility genes for lung diseases in the major histocompatibility complex revealed by lung expression quantitative trait loci analysis* (**Chapitre 4**). Ce projet comportait deux objectifs. Le premier était d'identifier des eQTL pulmonaires dans la région chromosomique du complexe majeur d'histocompatibilité (MHC). Le deuxième était d'identifier de nouveaux gènes de susceptibilité pour plusieurs phénotypes et maladies pulmonaires (asthme, cancer du poumon, fibrose kystique, fonction respiratoire, maladie pulmonaire interstitielle). Cette étude a permis d'identifier de

nouveaux gènes pour l'asthme (*AGPAT1*, *CDSN*), le cancer du poumon (*BTN3A2*, *ZFP57*), les fonctions respiratoires (*MICB*) et la maladie pulmonaire interstitielle (*AGPAT1*). Cette étude a été réalisée sous la supervision du Dr Yohan Bossé en collaboration avec les membres du consortium d'eQTL pulmonaire. Mon rôle dans cette étude fut la réalisation des analyses bio-informatiques, l'interprétation des résultats, et l'écriture du manuscrit. L'article a été soumis dans l'*European Respiratory Journal* le 9 novembre 2015 et a été accepté pour publication le 29 mars 2016.

Le troisième article s'intitule *Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis* (**Chapitre 5**). L'objectif était d'identifier de nouvelles régions chromosomiques impliquées dans le développement de la MPOC. Cette étude a identifié 22 régions d'intérêt; 15 précédemment associées avec les fonctions respiratoires et quatre nouvelles (*EEFSEC*, *DSP*, *MTCL1*, et *SFTPD*). Cette étude a été réalisée par l'*International COPD Genetics Consortium*, un regroupement de 26 cohortes provenant de huit pays différents. Ce projet a été réalisé sous la supervision du Dr Michael Cho du *Brigham and Women's Hospital* de Boston et l'auteur principal est Brian D. Hobbs provenant du même institut. Je suis troisième auteur sur 76 et mon rôle fut la réalisation d'analyses bio-informatiques, plus spécifiquement des eQTL pulmonaires. L'article a été soumis dans la revue *Nature Genetics* le 6 juillet 2016 et a été accepté pour publication le 23 novembre 2016.

Le quatrième et dernier article de cette thèse se nomme *Comprehensive genetic testing for alpha-1 antitrypsin deficiency in COPD patients* (**Chapitre 6**). Ce projet vise à vérifier la faisabilité et l'utilité du séquençage d'ADN comme outil de diagnostic pour le déficit en alpha-1 antitrypsine et à caractériser les variations génétiques du gène *SERPINA1* dans la population canadienne. Premièrement, le séquençage de l'ADN permet d'accélérer grandement le diagnostic de la maladie et ce, rapidement et à faible coût. Deuxièmement, nous avons identifié à partir de 400 sujets 19 variations génétiques, dont 15 mutations faux sens, dans les séquences codantes du gène *SERPINA1*. Parmi ces 19 variations, une mutation n'avait jamais été identifiée précédemment. Cette étude est une collaboration entre le Réseau Canadien de Recherche Respiratoire et l'étude *CanCOLD*. Mon rôle dans ce projet fut l'analyse et l'interprétation des données de séquençage et j'en suis l'auteur

principal. Nathalie Gaudreault a réalisé le séquençage de l'ADN et les Dr François Maltais, Wan Tan et Jean Bourbeau ont participé au recrutement des sujets et à la conception de l'étude. Le Dr Yohan Bossé a participé à la conception de l'étude, l'interprétation des résultats et l'écriture du manuscrit. L'article sera soumis pour publication dans les prochains mois.

CHAPITRE 1

Introduction

1.1 : Maladie Pulmonaire Obstructive Chronique

La maladie pulmonaire obstructive chronique (MPOC), aussi appelée broncho-pneumopathie chronique obstructive (BPCO), est un ensemble de maladies (bronchite chronique, emphysème) qui causent une obstruction en partie irréversible des voies respiratoires. Selon les dernières projections de l'Organisation Mondiale de la Santé (OMS), la MPOC deviendra la troisième cause de décès à l'échelle mondiale d'ici 2030[1]. En 2012 au Canada, les maladies chroniques des voies respiratoires inférieures (asthme, MPOC) en étaient la cinquième cause[2]. Les principaux symptômes de la MPOC sont une dyspnée marquée (essoufflement), une production importante de mucus et une toux chronique. La MPOC se développe majoritairement chez des fumeurs de longue date[3], mais seulement de 15 à 25% des fumeurs vont développer des symptômes d'obstruction respiratoire. Les seuls traitements actuellement disponibles sur le marché visent à soulager les symptômes et non à traiter la maladie.

Au cours des dernières décennies, des études effectuées sur des jumeaux, des familles et des individus non apparentés ont permis d'établir que la génétique jouait un rôle important dans le développement de la MPOC[4–7]. À ce jour, environ 200 gènes différents ont été associés avec cette pathologie, mais les résultats varient grandement d'une étude à l'autre[8]. Le seul gène connu à ce jour pouvant causer la MPOC sans tabagisme se nomme *SERPINA1* (*Serpin Peptidase Inhibitor, Clade A*). Des variations génétiques présentes dans la séquence codante de ce gène rendent la protéine non fonctionnelle entraînant un déficit en alpha-1 antitrypsine (AAT)[9].

Le développement de traitements permettant de résorber, stopper ou ralentir la progression de la MPOC est actuellement une priorité. Pour ce faire, il est primordial de mieux comprendre les mécanismes moléculaires impliqués dans le développement de cette pathologie potentiellement mortelle.

1.1.1 Historique de la MPOC

La MPOC et plus spécifiquement ses deux principales composantes, la bronchite chronique et l'emphysème, sont connus depuis plusieurs siècles[10]. La première mention d'une maladie ayant comme symptôme un essoufflement anormal remonte aux alentours de 400 av. J.-C.[11]. À cette époque très lointaine, Hippocrate aurait utilisé le terme asthme pour décrire une maladie caractérisée par une dyspnée, mais il ne faisait aucune distinction sur les causes de l'essoufflement. L'emphysème et la bronchite chronique ont fait leur apparition dans le vocabulaire médical plus de 2000 ans après Hippocrate et sa notion d'asthme.

L'emphysème fut décrit pour la première fois par le médecin suisse Théophile Bonet en 1679 dans son livre *Sepulcretum*[10,12] (Cimetière). À la suite d'environ 3000 autopsies effectuées par Bonet et ses collègues, l'emphysème a été décrit comme une condition médicale caractérisée par des poumons plus larges (*voluminous lung*). En 1769, l'italien Giovanni Battista Morgagni, un grand admirateur des travaux de Bonet, a publié son livre *De Sedibus et causis morborum per anatomem indagatis*[13,14] (Les sièges et les causes des maladies étudiées par l'anatomie) qui décrivait les poumons emphysémateux comme étant enflés et gonflés. En 1793, le pathologiste écossais Matthew Baillie publia *The Morbid Anatomy of Some of the Most Important Parts of the Human Body*[15] (L'anatomie morbide de certaines des parties les plus importantes du corps humain), le premier livre de pathologie écrit en anglais, dans lequel il a décrit des poumons avec de larges cavités où le tissu pulmonaire était détruit. En 1803, il publia un ouvrage d'illustrations pour compléter son précédent livre[16] et on y retrouve l'une des premières illustrations d'un poumon emphysémateux (**Figure 1-1**). En 1819, le médecin français René Laennec, inventeur du stéthoscope, publia le livre De l'auscultation médiate ou Traité du diagnostic des maladies des poumons et du cœur[17] dans lequel il décrivit plusieurs autopsies de patients emphysémateux. Par exemple, il a remarqué que l'emphysème était principalement présent dans les pointes et les parties antérieures des lobes et qu'un grand nombre d'alvéoles étaient de taille supérieure à la moyenne. Il a également remarqué qu'un poumon emphysémateux était gonflé et qu'il « ne s'affaisse nullement à l'ouverture de la poitrine, dans laquelle il est étroitement serré. »

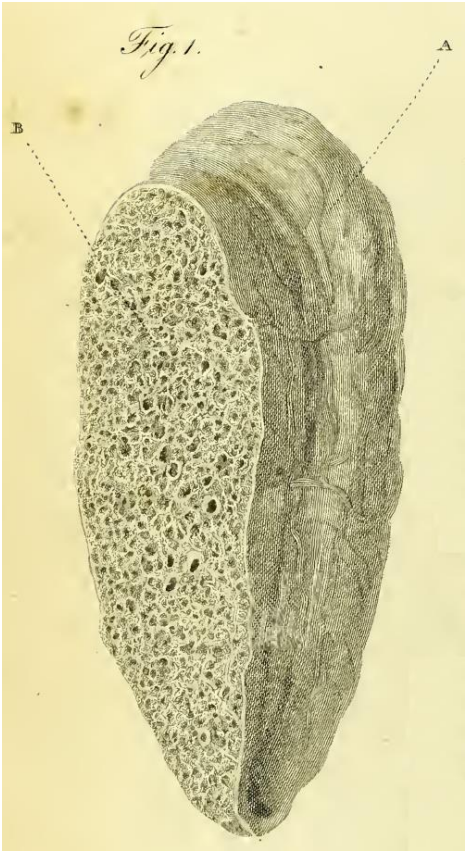

Figure 1-1 : Représentation graphique d'un poumon emphysémateux datant de 1803. Figure tirée de Baillie, M. *A series of engravings, accompanied with explanations, which are intended to illustrate the morbid anatomy of some of the most important parts of the human body.* 1803[16].

Le terme bronchite fut employé pour la première fois en 1814 dans le livre *An Essay on Bronchitis: With a Supplement Containing Remarks on Simple Pulmonary Abscess* (Un essai sur la bronchite: avec un supplément contenant des remarques sur l'abcès pulmonaire simple) écrit par le physicien anglais Charles Badham[18]. Celui-ci décrivit la bronchite comme une maladie inflammatoire des tissus sécréteurs présents dans les bronches respiratoires, qui résulte en une surproduction de mucus. En 1819, René Laennec distingua la bronchite chronique de l'emphysème[17]. Dans son ouvrage, la bronchite (catarrhe) chronique est caractérisée par « l'inflammation de la membrane interne ou muqueuse qui tapisse les bronches. » Il a également observé que la membrane des bronches est parfois

plus épaisse et d'une couleur rougeâtre. Le terme bronchite chronique (*chronic bronchitis*) a fait son apparition en 1837 dans l'ouvrage *A Treatise on the Diagnosis and Treatment of Diseases of the Chest: Diseases of the Lung and Windpipe* (Traité sur le diagnostic et le traitement des maladies du thorax: Maladies du poumon et de la trachée) du médecin irlandais William Strokes[19].

Suite aux observations effectuées entre la fin du 17e et le début du 19e siècle, plusieurs moments clés des 19e et 20e siècles ont permis de faire des bonds importants dans notre compréhension de la bronchite chronique et de l'emphysème. Parmi ces avancées se trouve l'invention du spiromètre en 1846 par John Hutchinson[20]. À ses débuts, cet appareil, qui est maintenant un incontournable du diagnostic de la MPOC, ne pouvait mesurer que la capacité vitale. Il faudra attendre jusqu'en 1947 pour voir apparaître les concepts de volume expiratoire maximal seconde (VEMS) et de capacité vitale forcée (CVF)[21]. Ces améliorations au spiromètre furent élaborées par Robert Tiffeneau et André Pinelli[22] et l'utilisation du coefficient de Tiffeneau, aussi appelé ratio VEMS/CVF et le rapport de Tiffeneau-Pinelli, a été grandement popularisé par Edward A. Gaensler[23,24]. Entre 1959 et 1965, trois rencontres importantes (*CIBA Guest Symposium*, 1959[25]; *American Thoracic Society Committee on Diagnostic Standards*, 1962[26]; *Committee on the Aetiology of Chronic Bronchitis*, 1965[27]) ont permis d'harmoniser les noms et les définitions employés[10,24]. Le terme maladie pulmonaire obstructive chronique (*chronic obstructive pulmonary disease*) fut employé pour la première fois par William Briscoe en 1965[28] et il est maintenant le nom officiel de la maladie. Les cinq décennies suivant la première mention du terme MPOC ont été le théâtre d'avancées rapides dans notre compréhension de la maladie.

1.1.2 Définitions et descriptions

La MPOC est une maladie caractérisée par une obstruction lente et progressive des voies respiratoires[29]. Toutefois, une étude récente a démontré qu'environ 50% des patients avec MPOC avaient déjà des fonctions respiratoires abaissées au début de l'âge adulte (Lange et al. NEJM 2015). Ces résultats suggérent que la MPOC chez ces patients n'est pas le résultat d'une perte progressive et accélélée des fonctions respiratoires, mais plutôt de l'incapacité d'atteindre des fonctions respiratoires normals à l'âge adulte. L'obstruction respiratoire observée dans la MPOC est en partie réversible lors de l'utilisation de bronchodilatateur[30,31]. Les deux principales composantes responsables de cette obstruction sont la bronchique chronique et l'emphysème, et celles-ci peuvent être présentes à différente proportion chez un même patient.

1.1.2.1 Bronchite chronique

La bronchite chronique se caractérise par une réduction de la lumière des bronches causée par l'inflammation des voies aériennes centrales[32–34] (**Figure 1-2**). Cette inflammation entraîne une surproduction de mucus, une réduction de la clairance muco-ciliaire et une augmentation de la perméabilité des cellules épithéliales bronchiques. La combinaison de tous ces processus fait en sorte que la lumière bronchique est réduite (**Figure 1-2B**) et ceci entraîne une augmentation de la résistance au passage de l'air[35]. Cette condition est diagnostiquée lors de la présence d'une toux persistante et d'une production d'expectoration pendant trois mois consécutifs et ce, durant deux années consécutives[36]. Selon le rapport de 2016 du *Global Initiative for Chronic Obstructive Lung Disease* (GOLD)[37], la bronchite chronique peut être présente chez des personnes avec des fonctions respiratoires normales.

1.1.2.2 Emphysème

L'emphysème est défini par une destruction irréversible des parois alvéolaires distales aux bronchioles terminales menant à la création de larges espaces aériens où les échanges gazeux sont affectés (**Figure 1-2C**). Deux types d'emphysème (centribulaire, panlobulaire) peuvent être observés chez les patients MPOC[38,39]. L'emphysème

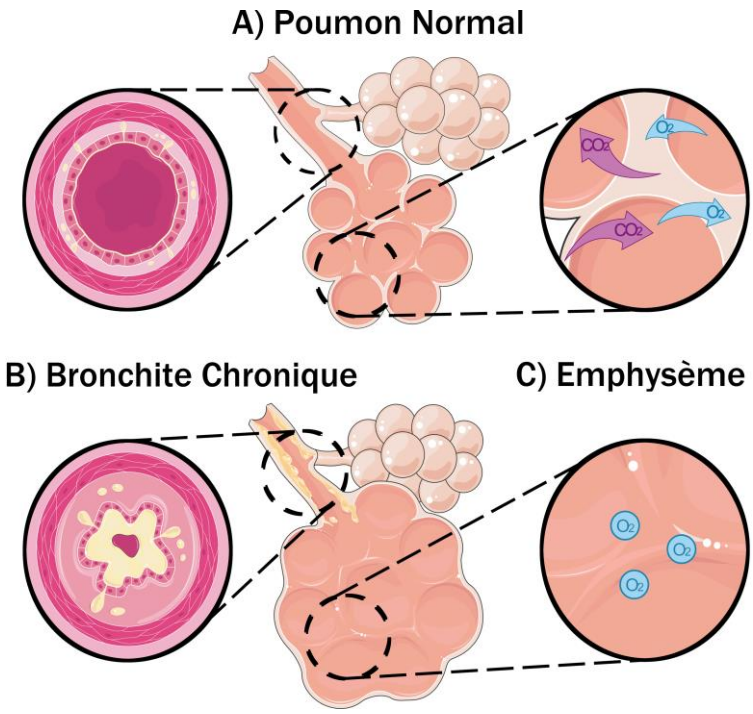

Figure 1-2 : Différences entre un poumon sain et un poumon de patient MPOC.
A) Bronche (gauche) et alvéoles (droite) d'une personne ne souffrant pas de MPOC. B) Une bronche enflammée et obstruée de mucus. C) Un espace mort où les échanges gazeux sont impossibles. Figure réalisée à l'aide des images de *SERVIER Medical Art.*

centribulaire est principalement développé chez les fumeurs de longue date[35,40]. Suite à une exposition prolongée à la fumée de cigarette, une inflammation se développe dans les alvéoles pulmonaires et provoque des dommages à leur structure [41]. Graduellement, les alvéoles se rompent et fusionnent pour créer de grandes poches d'air créant des zones ventilées mais peu ou pas perfusées aussi appellées espaces morts. L’espace mort, lorsque sévère, est responsable d’hypercapnie chez le patient MPOC. La destruction alvéolaire se produit principalement dans les lobes supérieurs et se concentre dans les portions centrales des lobules secondaires[40,42]. L'emphysème panlobulaire est pour sa part présent chez les patients atteints d'un déficit en AAT. L'absence de cette protéine provoque un débalancement de la balance protéase-antiprotéase. Les alvéoles présentes dans les lobes

inférieurs sont principalement affectées[40,42]. De plus amples détails sur le déficit en AAT sont présentés à la section 1.1.5.2.

1.1.3 Symptômes et diagnostic de la MPOC

La MPOC est une maladie complexe accompagnée de nombreux symptômes (dyspnée, toux, expectoration). Ceux-ci font leur apparition progressivement, mais leurs gravités peuvent augmenter drastiquement lors de crises aiguës (exacerbations). En plus des symptômes principaux, de nombreux troubles connexes (comorbidités) sont régulièrement observés chez ces patients. Le diagnostic représente un défi, car la population en générale connaît peu cette maladie, les symptômes sont communs et les professionnels de la santé n'ont pas nécessairement recours aux tests de spirométrie[43]. Les sections suivantes aborderont ces différents aspects cliniques en profondeur.

1.1.3.1 Diagnostic

Un diagnostic de MPOC doit être considéré chez les patients ayant une dyspnée, une toux chronique, une surproduction de mucus ou une exposition prolongée aux produits du tabac[37,44] (**Tableau 1-1**). Des tests de spirométrie post-bronchodilatateur (suivant l'inhalation d'un bronchodilatateur) doivent être effectués pour vérifier la présence d'obstruction respiratoire irréversible. Les deux principales variables mesurées lors de ces tests sont le Volume Expiratoire Maximal Seconde (VEMS) et la Capacité Vitale Forcée (CVF). Le VEMS est la quantité d'air expiré en une seconde lors d'une expiration maximale et la CVF est la quantité d'air totale expiré lors d'une expiration maximale. Lorsque le ratio entre le VEMS et la CVF est inférieur à 0.70, il y a présence d'obstruction respiratoire[37]. Le niveau d'obstruction est par la suite défini selon le VEMS (**Tableau 1-2**).

Tableau 1-1 : Symptômes cliniques à considérer pour le diagnostic de la MPOC.

Dyspnée	
	Persistante Progressive À l'effort
Toux chronique	
	Intermittente ou en continue Jour et/ou nuit
Expectoration	
	Production supérieure, peu importe la quantité et la fréquence
Exposition à des facteurs de risque	
	Tabagisme (primaire, secondaire) Fumée (cuisson, chauffage) Occupation professionnelle (poussière, produit chimique)
Autre	
	Historique familiale de MPOC

Tableau 1-2 : Classification des stades d'obstruction respiratoire.

Stade	**Critère du Volume Expiratoire Maximal Seconde**
GOLD 1 : MPOC légère	VEMS ≥ 80% de la valeur prédite
GOLD 2 : MPOC modérée	50% ≤ VEMS < 80% de la valeur prédite
GOLD 3 : MPOC sévère	30% ≤ VEMS < 50% de la valeur prédite
GOLD 4 : MPOC très sévère	VEMS < 30% de la valeur prédite

Le diagnostic de la MPOC n'est pas exclusivement basé sur les tests de spirométrie. Le médecin traitant a également recours à d'autres tests pour baser sa décision[37]:

- Imagerie médicale[42] : Permet de vérifier la présence de maladies ayant des symptômes semblables et d'identifier le sous-type d'emphysème (centribulaire, panlobulaire).
- Volume pulmonaire et capacité de diffusion : Aide à caractériser la sévérité de la maladie.
- Oxymétrie et relevé des gaz artériels du sang : Donne les niveaux sanguins d'oxygène et de dioxyde de carbone pour mieux adapter les soins.

- Dépistage du déficit en AAT[45] : Recommandé chez les patients plus jeune (<65 ans) ou ceux avec une faible exposition tabagique (<20 paquet année).

- Épreuve d'effort[46] : Excellent indicateur de la santé générale du patient.

1.1.3.2 Symptômes principaux

Le principal symptôme affectant les personnes atteintes de MPOC est une dyspnée qui se développe sur de nombreuses années[30,37,47] (**Figure 1-3**). L'essoufflement fait son apparition lors d'activités nécessitants un effort léger (e.g. monter les escaliers, tondre la pelouse). Par la suite, la progression de l'obstruction respiratoire rend les activités quotidiennes de plus en plus difficiles à exécuter. Dans les cas les plus avancés de la maladie, faire quelques pas sur une surface plane devient une tâche colossale[48]. La dyspnée est principalement causée par l'obstruction bronchique associée à une diminution du calibre des bronches (bronchite chronique) et/ou par la destruction du parenchyme pulmonaire (emphysème).

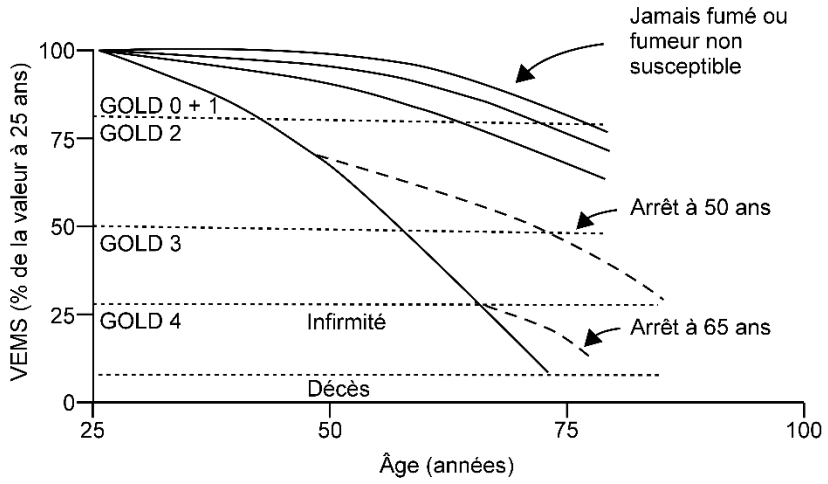

Figure 1-3 : Déclin du VEMS chez les fumeurs susceptibles.
Figure adaptée de J.C. Hogg, *Pathophysiology of airflow limitation in chronic obstructive pulmonary disease. Lancet* 2004; 364:709-721[35].

Un autre symptôme présent chez une forte proportion des patients MPOC, principalement ceux souffrant de bronchite chronique, est une toux persistante[35,37]. Ce symptôme est régulièrement le premier à faire son apparition, mais est souvent ignoré par les patients puisqu'il est souvent associé à l'utilisation du tabac[37]. La toux est présente en intermittence dans les stades légers de la maladie, mais fait son apparition quotidiennement dans les stades avancés. Bien que la toux soit l'un des principaux symptômes, son apparition ne signifie pas automatiquement la présence de MPOC. L'asthme, le cancer du poumon et le reflux gastro-œsophagien sont d'autres conditions pouvant occasionner une toux chronique[37].

Les patients MPOC peuvent produire de grandes quantités d'expectorations[37]. Lorsqu'il y a présence de pus dans le crachat, on parle alors d'expectorations purulentes et celles-ci sont associées avec la présence de médiateurs inflammatoires et/ou d'infection [49].

1.1.3.3 Exacerbations

Les personnes atteintes de MPOC peuvent vivre des périodes où leurs symptômes s'aggravent rapidement (exacerbations)[50]. Les principales causes de ces crises aiguës sont les infections bactériennes et/ou virales (50%)[51–55] et la pollution atmosphérique (15 à 20%)[50,56,57]. Les exacerbations sont responsables de 25% des pertes totales de fonctions respiratoires observées[50,58,59]. La fréquence des exacerbations augmente graduellement avec le stade de la maladie[50]. Selon une étude publiée par Hurst *et al.* en 2010[50], la fréquence des crises aiguës par année serait de 0.85 pour les GOLD 2, de 1.34 pour les GOLD3 et de 2.00 pour les GOLD4. Les patients souffrants d'exacerbations récurrentes sont plus souvent atteints de reflux gastro-œsophagien[50,60–62].

1.1.3.4 Comorbidités

Les personnes atteintes de MPOC sont régulièrement atteintes de comorbidités qui ont pour conséquences une aggravation de leur condition et une détérioration de leur qualité de vie[63–66]. En 2013, Vanfleteren *et al.*[67] ont remarqué que la grande majorité de leurs patients MPOC modérés, sévères et très sévères (97.7%) souffraient d'une ou de plusieurs

comorbidités, et que 53.5% des sujets vivaient avec quatre comorbidités et plus. En 2009, Barr *et al.*[68] ont réalisé une enquête auprès de 1003 sujets MPOC. Le nombre médian de comorbidités observées était de neuf et un plus grand nombre de comorbidités était associé à une dyspnée plus sévère. Les comorbidités observées sont nombreuses (anémie, anxiété, asthme, cancer du poumon, dépression, diabète, dysfonction musculaire, infection respiratoire, maladies cardiovasculaires, maladies rénales, ostéoporose, reflux gastro-œsophagien, syndrome métabolique, etc.) et affectent différents organes du corps humain[69] (**Figure 1-4**; **Tableau 1-3**).

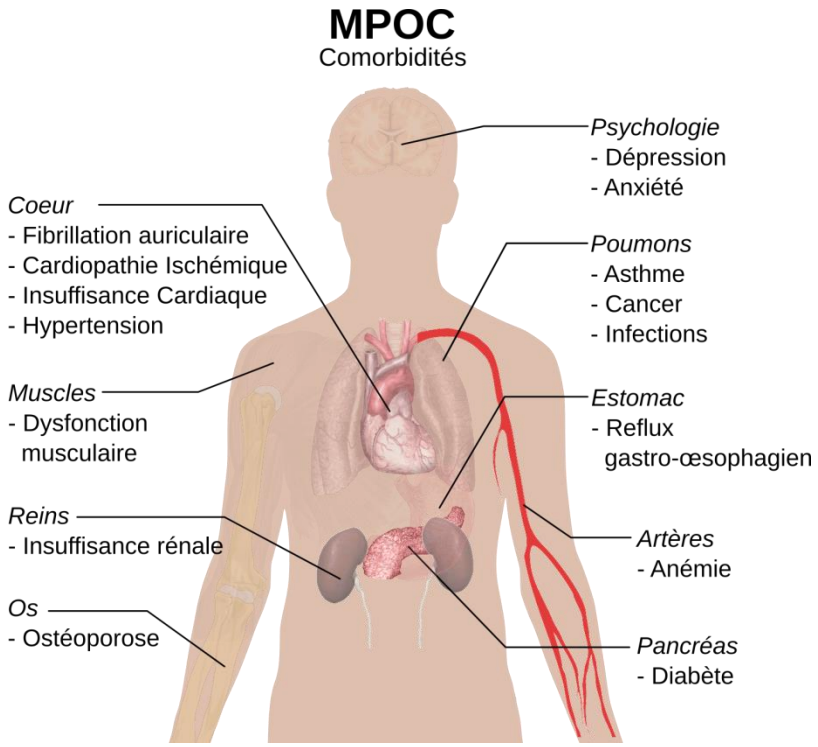

Figure 1-4 : Comorbidités associées à la MPOC.

Figure réalisée grâce aux images de la *Wikipedia foundation: Human body diagrams*.

Tableau 1-3 : Caractéristiques des comorbidités associées à la MPOC.

Anémie	
Prévalence	7.5 - 34%[69,70]
Impact	Qualité de vie réduite[71] Plus grand nombre d'hospitalisations[72] Mortalité accrue[73]
Cause	Prévalence élevée de maladies rénales chez les patients MPOC[74]
Anxiété et dépression	
Prévalence	Dépression : Entre 10 et 42% des sujets MPOC[75–77] et jusqu'à 60% chez les patients sous oxygénothérapie[78] Anxiété : Entre 7 et 50% des patients[79]
Impact	Exacerbations[80] et hospitalisations[81] plus fréquentes Perte d'autonomie[82] Qualité de vie réduite[83] Mortalité accrue[84]
Cause	Sévérité de la maladie[85,86]
Cancer du poumon	
Prévalence	25%[87]
Impact	Mortalité accrue[88,89]
Cause	Prédisposition génétique commune avec la MPOC (e.g. *CHRNA3/5*)[90–95] Inflammation chronique[96–98] Stress oxydant[99]
Dysfonction musculaire	
Prévalence	33%[100]
Impact	Mortalité accrue[101] Capacité d'exercice diminuée[100] Qualité de vie réduite[100]
Cause	Déficience en vitamine D[102,103] Inflammation[104,105] Stress oxydant[106,107] Hypoxémie[108] Hypercapnie[109,110]
Infection respiratoire	
Impact	Exacerbations[53,54]
Cause	Mucus (hypersécrétion, viscosité élevée, évacuation réduite)[111] Barrière épithéliale des bronches et des poumons endommagée[112]
Maladies cardiovasculaires	
Prévalence	Fibrillation auriculaire : 11-14%[113,114] Hypertension artérielle : 40-60%[68,69]

	Insuffisance cardiaque : 5-7%[69] Maladie coronarienne : 28-70%[115]
Impact	Insuffisance cardiaque : Survie réduite lorsque hospitalisé[116] Maladie coronarienne : Qualité de vie et capacité à l'effort réduite, dyspnée plus sévère et exacerbations de plus longue durée[117]
Cause	Fibrillation auriculaire : Hypoxie[118,119] Hypertension artérielle : Rigidité artérielle augmentée[120] Maladie coronarienne : Inflammation systémique, stress oxydant et hypoxémie[121]
Maladies rénales	
Prévalence	Entre 1 et 11%, dépendant du type de maladie rénale[69,122]
Impact	Mortalité accrue[123]
Cause	Inflammation systémique[124] Hypoxie[125]
Ostéoporose	
Prévalence	Entre 9 et 69% des patients MPOC[126,127]
Impact	Exacerbations et hospitalisations plus fréquentes[128]
Cause	Inflammation systémique[126] Déficience en vitamine D Utilisation de corticostéroïdes
Reflux gastro-œsophagien	
Prévalence	50%[129,130]
Impact	Plus grand nombre d’exacerbations[62] Risque accru d'admission aux soins intensifs[131]
Cause	Perte de sensibilité laryngo-pharyngée[132] Pression intra-abdominale élevée[133] Pression intra-thoracique négative[133] Utilisation de certains médicaments (théophylline, β2-adrénergique)[134]
Syndrome métabolique et diabète	
Prévalence	Diabète : 12%[77,135] Syndrome métabolique : 20% des sujets non-emphysémateux et 10% des patients emphysémateux[135]
Impact	Fonctions respiratoires plus faibles[136] Qualité de vie réduite[136] Nombre d'hospitalisation accru[137] Mortalité supérieure[138,139]
Cause	Inflammation systémique[135]

L’asthme est régulièrement présente chez les patients MPOC[140]. Celui-ci se caractérise par une inflammation des grandes et petites voies respiratoires. Il se manifeste généralement à l’enfance et est régulièrement accompagné d’allergies. Tout comme les sujets MPOC, les asthmatiques ont un VEMS réduit, mais celui-ci retourne à la normale suivant l’utilisation d’un bronchodilatateur. Chez les patients plus âgés, une baisse permanente du VEMS peut s’installer. L’asthme est une comorbidité particulière pour la MPOC puisque la communauté médicale s’intéresse de plus en plus à un phénomène qui se nomme syndrome de chevauchement de l’asthme et de la MPOC (*asthma–COPD overlap syndrome*)[141,142]. Ce terme désigne les individus qui possèdent des symptômes des deux maladies (e.g. dyspnée, fonctions respiratoires réduites, reflux gastro-œsophagien, toux chronique). Ce syndrome, qui n’est pas officiellement reconnu comme une maladie, fait preuve d’un grand engouement dans la communauté scientifique. Preuve de cette popularité grandissante, PubMed recense 1 article scientifique en 2011 pour les termes *asthma–COPD overlap syndrome*. Ce nombre grimpe à 2 en 2013, 8 en 2014, 44 en 2015 et 69 en 2016. Avec les données actuellement disponibles, il est difficile de savoir si l’asthme et la MPOC sont deux maladies distinctes avec des symptômes semblables ou bien une seule et unique maladie avec différentes manifestations cliniques. Cette situation découle partiellement des critères de sélection employés lors des études scientifiques des dernières décennies. La majorité des études sur l’asthme ou la MPOC excluaient les patients souffrant également de l’autre maladie[141,142]. Pour cette raison, il est difficile d’établir la prévalence de l’asthme chez les patients MPOC. Il est cependant reconnu que les sujets asthmatiques possèdent un risque supérieur de développer la MPOC[143]. De nouvelles études sont donc nécessaires pour mieux caractériser ce syndrome et déterminer si l’asthme et la MPOC sont une ou deux maladies.

1.1.4 Prévalence et impact économique

La MPOC est une maladie dont la prévalence augmente avec l'âge et se développe habituellement chez les fumeurs de longue date[144]. Les symptômes de la maladie apparaissent rarement avant l'âge de 55 ans. Selon une méta-analyse de 123 études effectuée en 2015, la prévalence mondiale estimée de MPOC était de 10,7% (227.3 millions) en 1990 et de 11,7% (384 millions) en 2010[145]. Au Canada, plus de 800,000

personnes souffraient de MPOC en 2014[146] et le quart de ce nombre se situait dans la province de Québec.

La prévalence de MPOC varie grandement selon le sexe. Au niveau mondial, les hommes sont plus affectés que les femmes[145]. Par contre, cette tendance semble s'inverser dans certains pays industrialisés, dont le Canada et les États-Unis[147]. Au Canada en 2014, les femmes (445,637) étaient majoritaires comparativement aux hommes (358,406)[146]. Au Québec en 2011, 31,318 hommes et 15,955 femmes ont été diagnostiqués avec la maladie et le total d'hommes et de femmes affectés en 2011 était de 163,128 et 160,652 respectivement[147]. Aux États-Unis en 2011, 7,0% des femmes âgées de 25 ans et plus souffraient de MPOC[75], tandis que les hommes étaient atteints dans une proportion de 4,3%. Au cours de la dernière décennie, plusieurs études ont constaté que la mortalité occasionnée par la MPOC diminuait chez les hommes, mais augmentait ou demeurait stable chez les femmes[75,148]. La prévalence de la maladie varie également selon l'âge[37].

La proportion de personnes souffrant de MPOC augmente avec le vieillissement de la population (**Figure 1-5**). Par contre, les connaissances actuelles ne permettent pas de déterminer si le vieillissement augmente les chances de développer la maladie ou ne fait que représenter l'accumulation des dommages aux poumons[37].

En plus d'encombrer nos systèmes de santé, la MPOC a également des impacts financiers important pour notre société. Entre 2001 et 2010 en Colombie-Britannique au Canada, il en coûtait en moyenne 8,600$ par année pour traiter une personne MPOC[149]. L'hospitalisation des patients représentait la majorité des coûts engendrés et le nombre d'admission par patient augmentait avec la progression de la maladie. En 2008, le fardeau économique des hospitalisations uniquement a été estimé à 1,5 milliard de dollars au Canada[150]. De plus, les coûts directs et indirects engendrés par la maladie devraient grimper jusqu'à 9 milliards d'ici 2030[151]. Aux États-Unis en 2010, le fardeau économique a été de 36 milliards de dollars américains et ce montant grimpera à 49 milliards en 2020[152].

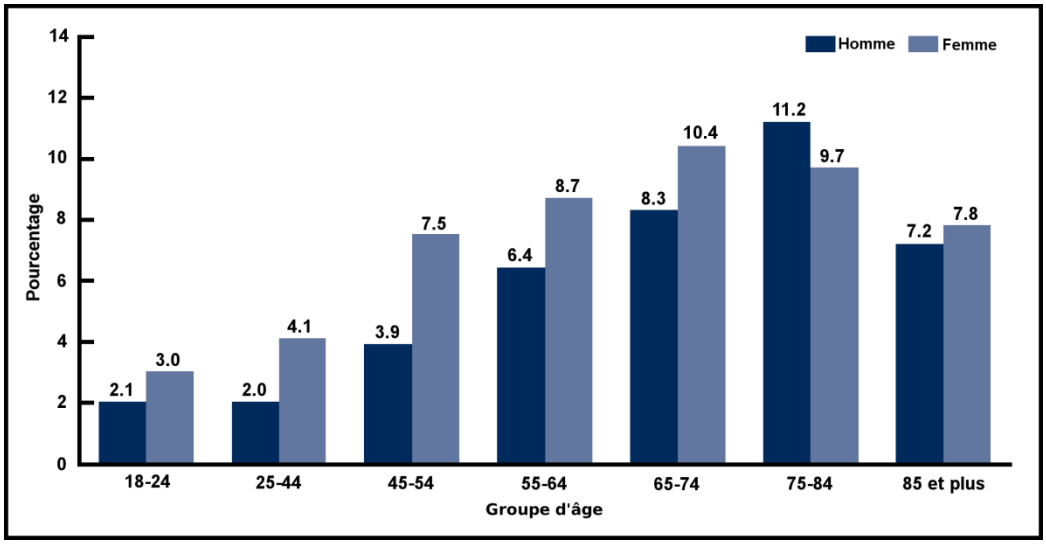

Figure 1-5 : Prévalence de la MPOC selon le groupe d'âge et le sexe aux États-Unis entre 2007 et 2009.

Figure adaptée du *Centers for Disease Control and Prevention: National Center for Health Statistics*.

1.1.5 Causes

Tel que mentionné précédemment, le tabagisme est le principal facteur de risque pour la MPOC. L'utilisation des produits du tabac est responsable d'environ 90% des décès chez les patients MPOC[153–155]. La proportion de fumeurs souffrant d'obstruction respiratoire varie grandement d'une étude à l'autre. Les estimations les plus courantes font état de 15 à 25% de fumeurs affectés[156], mais certaines études avaient jusqu'à 50% de fumeurs affligés[157,158]. Ces écarts peuvent s'expliquer par des méthodologies différentes[159] et l'évolution des techniques employées pour le diagnostic[3]. Bien que la majorité des cas de MPOC soit causés par le tabagisme, il existe une grande variabilité parmi ceux-ci. Le stade (GOLD 1 à 4) et le phénotype (emphysème et/ou bronchite chronique) varient grandement entre les patients. Certaines prédispositions génétiques semblent donc jouer un rôle dans le développement et l'évolution de la maladie[8,160].

1.1.5.1 Tabagisme

Le tabagisme est de loin le principal facteur conduisant au développement de la MPOC. La fumée de cigarette, avec plus de 4000 agents chimiques[161], a des effets néfastes autant sur les voies biologiques (immunité innée[162], stress oxydant[163,164]) que sur la structure du poumon (destruction des cils vibratiles[165]). L'accumulation de ces changements aura pour conséquence une réduction des fonctions respiratoires chez les personnes susceptibles.

1.1.5.1.1 Immunité innée

L'immunité innée est constituée d'un nombre important de cellules (éosinophiles, neutrophiles, macrophages, cellules *NK*) constituant la première ligne de défense du système immunitaire. Ces cellules ont comme fonction d'éliminer les agents infectieux[166], de se débarrasser des cellules mortes et de réparer les dommages causés au tissu pulmonaire[167]. La trop grande présence de neutrophiles et de macrophages dans les poumons entraîne un déséquilibre de la balance protéase-antiprotéase et un stress oxydant chronique.

Déséquilibre protéase-antiprotéase

Les protéases, produites en partie par les neutrophiles et les macrophages[168], sont impliquées dans la migration cellulaire[169], la défense contre les pathogènes[170] et la réparation cellulaire[171]. Dans les poumons, l'élastase des neutrophiles est un antibactérien puissant qui s'attaque à la matrice extracellulaire des pathogènes[172,173]. Lorsque l'enzyme est présente en grande quantité, le tissu sain est dégradé et de l'emphysème se développe[168,174,175]. L'élastase des neutrophiles endommage l'élastine et le tissu conjonctif des poumons[174,175]. Ces dommages provoquent l'apoptose des cellules épithéliales et endothéliales pulmonaires et résultent en un élargissement des alvéoles. Les antiprotéases ont pour fonction d'inhiber les protéases et ainsi protéger le tissu sain. Un déséquilibre de ce système est présent chez certains fumeurs souffrant de MPOC[168] et chez les individus atteints de déficit en alpha-1 antitrypsine[176].

Stress oxydant

Le stress oxydant est caractérisé par la présence d'une grande quantité de dérivés réactifs de l'oxygène, mieux connu sous le nom de *reactive oxygen species* (ROS), dans le poumon. Les ROS ont comme principales sources la fumée de cigarette et la poussée oxydative des cellules inflammatoires (*respiratory burst*)[177,178]. Dans des conditions normales, les ROS sont impliqués dans plusieurs mécanismes, dont la défense immunitaire, l'apoptose et la signalisation cellulaire[179–181]. Par contre, une trop grande présence de ROS provoque des dommages à la structure des cellules, à l'ADN et aux protéines[178,180,182] et induisent la production de médiateurs inflammatoires (e.g. NF-κB, TNF, IL-8)[164,178]. Le tabagisme est la principale source de ROS dans les premiers stades de la maladie, mais le stress oxydant persiste suite à l'arrêt tabagique[183]. Les cellules inflammatoires pulmonaires (neutrophiles, macrophages), qui sont présentes en grand nombre chez les patients MPOC, relâchent une grande quantité de ROS. Le *respiratory burst* débute pendant l'utilisation des produits du tabac et se poursuit après l'arrêt tabagique[184,185]. Ce phénomène s'explique par la réponse immunitaire adaptative qui s'installe chez certains individus susceptibles.

1.1.5.1.2 Immunité adaptive

La réponse immunitaire adaptative est en partie responsable des symptômes persistants suite à l'arrêt tabagique (e.g. dommage cellulaire, inflammation) [3,186,187]. Normalement, la réaction immunitaire se termine lorsque l'agression prend fin (e.g. infection, tabagisme), mais ce n'est pas le cas chez les patients MPOC. Les mécanismes responsables de ce phénomène ne sont pas connus avec certitude, mais certaines molécules semblent impliquées[188]. Plusieurs pistes sont actuellement étudiées pour expliquer leur activation continue : présence de pathogènes, réaction à la fumée de cigarette, développement d'une réaction auto-immunitaire[189].

Lymphocytes B

Les lymphocytes B ont pour fonction la production d'anticorps et de cytokines[190]. Ceux-ci sont produits dans la moelle osseuse et migrent ensuite vers la rate et les ganglions[191]. Lors d'une infection, ces cellules participent à l'identification de l'agent infectieux et à la régulation de la réponse immunitaire. Chez les patients MPOC, la quantité de lymphocytes B dans les voies respiratoires augmente avec la progression de la maladie[192]. Ceux-ci forment des follicules lymphoïdes avec les lymphocytes T CD4+ dans le parenchyme et les parois bronchiques des patients MPOC[189,192,193]. L'activation continue des lymphocytes B explique en partie l'inflammation chronique présente chez les patients MPOC.

Lymphocytes T CD4+

Les lymphocytes T CD4+ sont des cellules impliquées dans la régulation de la réponse immunitaire. Ceux-ci accomplissent cette tâche en aidant les lymphocytes B à produire des anticorps, en activant les macrophages, et en recrutant des neutrophiles, éosinophiles et basophiles au site d'infection et d'inflammation[194–197]. Afin d'accomplir toutes ces tâches et de produire une réponse adaptée, les cellules CD4+ naïves doivent subir plusieurs étapes de transformations pour se spécialiser. Cette différentiation est affectée par plusieurs facteurs[197] (e.g. cytokines présentes dans le milieu de maturation, durée de l'exposition) et vise à produire des lymphocytes T auxiliaires, mieux connus sous le nom de *T helper cells* ou *Th cells*[197]. Par exemple lorsque les lymphocytes CD4+ naïfs sont exposés à de grandes quantités d'IL-12 et d'interféron gamma (IFNγ)[198], ceux-ci se différentient en lymphocyte Th1, des cellules T impliquées dans l'inflammation et l'activation des macrophages. À la fin de la réponse immunitaire, un faible nombre se transforment en lymphocyte T à mémoire (*memory T cell)*. Chez les patients MPOC, il y a présence de lymphocytes CD4+ dans les voies respiratoires et dans le parenchyme pulmonaire[192,199–203]. De plus, la quantité de lymphocytes augmente avec la progression de la maladie. La présence continuelle de ces lymphocytes explique en partie l'inflammation chronique et les dommages aux tissus pulmonaires observés chez ces individus[200].

Lymphocytes T CD8+

Les lymphocytes T CD8+ sont grandement impliqués dans la réponse immunitaire contre les infections. Lorsqu'ils sont activés, ces lymphocytes vont éliminer les cellules infectées grâce à trois mécanismes différents. Le premier est la sécrétion de cytokines antibactériennes telle que TNF-α et IFN-γ[204]. Le deuxième est la production et le relâchement de granules cytotoxiques, principalement des perforines et des granzymes[204,205]. Finalement, les lymphocytes CD8+ peuvent détruire les cellules infectées via l'interaction entre leur ligand FasL et le récepteur FasR[205,206]. Tout comme les lymphocytes T CD4+, les CD8+ sont présents dans les poumons et les voies respiratoires des patients MPOC et leur présence augmente également avec la gravité de la maladie[207–215]. Cette présence continue contribue à la détérioration du parenchyme pulmonaire et à l'inflammation chronique des voies respiratoires[186].

1.1.5.2 Déficit en alpha-1 antitrypsine

L'apha-1 antitrypsine est une antiprotéase produite par le foie responsable d'inhiber l'élastase neutrophile et d'autres protéases[216]. Un déficit en apha-1 antitrypsine (DAAT) peut entraîner le développement d'emphysème chez les non-fumeurs et causer un emphysème précoce chez les fumeurs. Le DAAT est causé par une ou des mutations dans la séquence codante du gène *SERPINA1* qui est localisé sur le chromosome 14. En Amérique du Nord, il est estimé qu'une personne sur 3,000 à 5,500 souffre de cette maladie génétique[217]. Chez les patients MPOC, entre 1 et 5% de tous les cas en sont le résultat[218–220]. La proportion de gens affectés tant dans la population générale que chez les patients MPOC est probablement sous-estimée[221].

Chez les individus soupçonner d'être atteint de DAAT (sujet d'environ 40 ans, tabagisme moindre), plusieurs techniques sont actuellement employées[219]: mesure de la quantité de protéine présente dans le sérum ou le plasma sanguin, caractérisation de la protéine par focalisation isoélectrique, génotypage des mutations les plus courantes, séquençage des régions codantes du gène *SERPINA1*. La quantité normale de protéine dans le sang est de 1.5 à 3.5 g/L (150 à 350 mg/dL)[220]. Une personne souffrant de DAAT aura des niveaux de protéines inférieurs à 0.8 g/L (80 mg/dL)[220]. La quantification de la

protéine est la première étape à effectuer lorsqu'un médecin soupçonne une déficience[222,223] et plusieurs techniques existent : immunodiffusion radiale, néphélémétrie, et dosage immunoturbidimétrique renforcé de latex [224]. Une des limitations de ce test est que les niveaux d'AAT peuvent varier lors d'épisodes inflammatoires et d'infections[225,226]. Tel que mentionné précédemment, les patients MPOC ont un risque élevé de développer des infections. Les médecins traitants doivent donc être prudents lors de l'interprétation des résultats de ce test. Suite à la quantification sanguine, la caractérisation de la protéine et le génotypage des deux variants de risque les plus courants sont effectués (Pi Z et S). La caractérisation de la protéine s'effectue par focalisation isoélectrique. Celle-ci permet de détecter si le sujet est porteur d'un des quatre variants analysés (F, M, S ou Z), dépendamment de la vitesse de migration de la protéine. Cette technique comporte également plusieurs limitations : nombre limité de variant testé, doit corréler avec les niveaux sanguins de protéine et résultats difficiles à interpréter. Le génotypage des variants Pi Z et S consiste à déterminer les génotypes des variations génétiques associées à ces variants[224]. Bien qu'il existe un grand nombre de variants pouvant occasionner une déficience, seulement les Pi Z et S sont normalement génotypés puisqu'ils sont responsables de la grande majorité des cas. Lorsqu'un DAAT est soupçonnée mais que les résultats de la caractérisation et du génotypage ne concordent pas avec les niveaux mesurés, le séquençage des régions codantes (exons) du gène *SERPINA1* est requis[219,227]. Contrairement aux techniques précédentes, le séquençage des exons permet de déterminer toutes les variations génétiques pouvant modifier la séquence de la protéine. Une fois le séquençage complété, il est possible de déterminer si une ou des mutations présentes dans la séquence codante du gène peuvent expliquer les niveaux d'AAT mesurés. Suite aux différentes analyses effectuées, le médecin traitant, avec l'aide d'un spécialiste en génétique médicale, peut déterminer si le patient souffre d'un DAAT ou non. Les différents aspects génétiques de la DAAT sont décrits en profondeur à la section 1.2.2.1.

1.1.5.3 Autres causes

Outre le tabagisme, d'autres facteurs de risques sont associés à un déclin rapide des fonctions respiratoires[228,229] : exposition à des particules fines (outre tabagisme), génétique, poids à la naissance, sexe, et statut socio-économique.

Exposition à des particules

La présence de particules en suspension dans l'air a été associée à une diminution des capacités respiratoires[230–233]. Les principales sources de particules sont les activités professionnelles, la pollution atmosphérique, et la combustion de biomasse pour la cuisson et le chauffage. Plusieurs molécules relâchées par ces sources (e.g. NO_2, O_3) ont des effets semblables à ceux du tabagisme (e.g. stress oxydant, inflammation pulmonaire)[232].

Génétique

Outre le DAAT occasionnée par le gène *SERPINA1*, d'autres gènes sont suspectés d'avoir un rôle dans le développement de la maladie. La section 1.2.2 décrit en profondeur la génétique de la maladie.

Poids à la naissance

Selon une étude de *Barker et al.* parue en 1991[234], un poids inférieur à la naissance est associé à des fonctions respiratoires réduites à l'âge adulte. Le risque de mourir de MPOC est également associé à un poids inférieur à la naissance et à un an. Une étude publiée par Stern *et al.* en 2007[235] a démontré que les jeunes adultes qui avaient un poids inférieur à la naissance ont des fonctions respiratoires réduites à 22 ans. D'autres études ont également démontré cette relation[236,237].

Sexe

Comme énoncé précédemment, dans la très grande majorité des pays, les hommes sont plus atteints de MPOC que les femmes, mais les femmes le sont maintenant autant, sinon plus affectées que les hommes dans certains pays industrialisés[238]. La

prédominance des hommes est directement liée à la plus grande utilisation des produits du tabac par ceux-ci. Selon les données de l'OMS, 34.4% des hommes contre 11.7% des femmes âgés de 15 ans et plus consommaient régulièrement des produits du tabac en 2015[239]. Des différences entre les sexes existent également dans la manifestation de la maladie[240]. L'emphysème serait plus fréquent chez les hommes et la bronchite chronique chez les femmes[241]. La dyspnée serait plus sévère chez les femmes malgré un nombre de paquet-année inférieur[242]. La dépression et l'anxiété affecteraient une plus grande proportion de femmes[238]. Les raisons précises expliquant ces observations sont toujours inconnues.

Statut socio-économique

Le statut socio-économique est principalement déterminé par le revenu et le niveau de scolarité d'une personne. Le risque de souffrir et de mourir de MPOC est supérieur chez les personnes avec un statut socio-économique[243–245] et un revenu moindre[245,246]. La pauvreté est également associée à de plus faibles fonctions respiratoire[247]. Cette situation s'explique par plusieurs facteurs associés à un statut moins élevé[229] : tabagisme supérieur, infection en bas âge, emploi.

1.1.6 Traitements

À ce jour, il n'existe aucune solution pour guérir la MPOC. Les traitements pharmacologiques et non pharmacologiques actuels sont employés pour diminuer les symptômes, diminuer la fréquence et la gravité des exacerbations, et améliorer la qualité de vie et la capacité physique des patients[37]. Peu importe le traitement employé, l'arrêt tabagique est l'élément central de toute intervention[88].

1.1.6.1 Traitements non pharmacologiques

Le traitement non pharmacologique le plus important est la cessation tabagique complète chez les fumeurs. Tel que mentionné à la section 1.1.5, le tabagisme est le principal facteur de risque pouvant occasionner la maladie. L'arrêt tabagique a comme principal effet de ralentir la progression de la maladie en réduisant l'apport de produits

néfastes dans les poumons[88] et cela se traduit par une diminution de la mortalité chez les ex-fumeurs comparativement aux fumeurs actifs[88,248,249].

Les infections respiratoires sont plus fréquentes chez les sujets MPOC et elles peuvent avoir de graves conséquences (e.g. exacerbation, décès)[250]. Celles-ci sont responsables de la moitié des exacerbations. [251,252]. L'influenza est détecté chez 28% des patients exacerbés et certaines études ont démontré que les patients MPOC sont plus susceptibles aux infections par ce virus[253]. La vaccination antigrippale est donc extrêmement importante pour diminuer les risques d'infections et autres complications associées.

La réadaptation pulmonaire est conçue pour réduire les symptômes, augmenter la qualité de vie et améliorer l'état de santé général des patients[254–259]. Les programmes de réadaptation pulmonaire sont composés d'exercices physiques, de séances d'éducation, de programmes alimentaires et de soutien psychologique. En plus des bienfaits pour le patient, ces programmes permettent de réduire les coûts pour le système de santé[260,261].

L’oxygénothérapie est parfois employée chez les patients MPOC sévères[262,263]. Ces patients sont caractérisés par une hypoxémie sévère au repos. Ceux-ci voient leur mortalité diminuée lorsque ce traitement est employé[264]. L’oxygénothérapie n’a toutefois aucun effet chez les patients dont la maladie est contrôlée et sans hypoxémie[265].

Chez les cas sévères de la maladie, des techniques plus invasives peuvent être employées (bullectomie, réduction du volume pulmonaire, transplantation pulmonaire)[3]. Celles-ci permettent de réduire la dyspnée[266], la mortalité[267] et d’améliorer la qualité de vie[267]. Par contre, ces techniques ne peuvent pas être employées chez tous les patients et requièrent des installations de pointes.

1.1.6.2 Traitements pharmacologiques

Les traitements actuellement disponibles ont pour but de soulager les symptômes de la maladie[37]. Il existe une gamme de médicaments qui sont prescrits selon la sévérité de la maladie et la réponse du patient.

Pour les cas de MPOC léger, la première classe de médicaments prescrits sont les bronchodilatateurs à courte durée d'action (BDCA). Cette classe de médicaments est recommandée lors de difficultés respiratoires passagères pour détendre les bronches. Lorsque les BDCA s'avèrent insuffisants, une combinaison d'anticholinergique à longue durée d'action (ACLA) et de BDCA ou une combinaison de β2-adrénergiques à longue durée d'action (BALA) et de BDCA peuvent être utilisée. Ces médicaments ont comme fonction de prévenir les bronchospasmes à court et à long terme.

Pour les cas plus sévères de la maladie, les traitements vont dépendre dans un premier temps de la fréquence des exacerbations aiguës et dans un deuxième temps de l'historique des médicaments utilisés. Les médicaments précédents seront utilisés selon des combinaisons et des doses différentes. Les corticostéroïdes en inhalation en association avec un β2-adrénergique à longue durée d'action (CSI/BALA) et la théophylline peuvent aussi être employés.

1.1.6.3 Traitement du déficit en alpha-1 antitrypsine

À l'heure actuelle, les médecins peuvent prescrire un traitement substitutif de l'AAT[45]. Cette thérapie vise à augmenter la concentration d'AAT dans les poumons à l'aide d'injections. Selon les lignes directrices de la Société Canadienne de Thoracologie, la thérapie devrait être prescrite chez les patients non-fumeurs et les anciens fumeurs avec une MPOC contrôlée[45]. Ces injections sont associées à une réduction de la mortalité et un ralentissement du déclin des fonctions respiratoires et de la densité pulmonaire[268,269]. Par contre, la fréquence des exacerbations et la qualité de vie des patients demeurent inchangées. De plus, les études précédentes n'ont pas mesuré plusieurs critères cliniques importants telle la dyspnée, la capacité physique et la fréquence d'hospitalisations. De nouvelles études cliniques devront être réalisées pour mieux documenter les effets à court et long terme de cette thérapie.

1.2 : Génétique de la maladie pulmonaire obstructive chronique

Selon le dictionnaire Larousse, la génétique est la « partie de la biologie qui étudie les lois de l'hérédité ». Plus particulièrement, ce domaine scientifique étudie le rôle des gènes dans différents phénomènes biologiques, notamment leur influence sur la santé et les maladies. La génétique a connu un essor fulgurant dans les dernières décennies et plus spécifiquement depuis la conclusion du *Human Genome Project* en 2003[270]. Un exemple flagrant de l'évolution des techniques est la diminution du coût et du temps de séquençage pour un génome humain. Le séquençage du premier génome humain a nécessité 13 années et a coûté 2,7 milliards de dollars américains (https://www.genome.gov). Aujourd'hui, le séquenceur *HiSeq X Ten System* de la compagnie *Illumina* (San Diego, Californie, États-Unis) peut séquencer 18,000 génomes humains par année pour un prix moyen de 1000$ par génome (http://www.illumina.com). Ces avancées technologiques ont permis de faire des progrès importants dans la compréhension de maladies complexes tel la MPOC.

Au cours des dernières décennies, la recherche génétique sur la MPOC a fait d'importants progrès. Outre le gène *SERPINA1*, plus de 200 gènes ont été associés avec la maladie[8] et ceux-ci sont impliqués dans une vaste gamme de fonctions biologiques tel la neurotransmission (*CHRNA3*), l'inflammation (*IL4*, *IL6*, *IL13*, *IL1B*, *IL1RN*, *LTA*, *TNF*, *TGFB1*) et l'apoptose (*CASP8*, *CASP10*). Par contre, les associations détectées ne font qu'augmenter ou diminuer minimalement le risque de développer la maladie. Les prochaines sections développeront sur l'état actuel de nos connaissances sur la génétique de la MPOC.

1.2.1 Introduction à la génétique

La section 1.2.1 présentera différentes notions de base de génétique humaine afin de faciliter la compréhension de cette thèse.

1.2.1.1 Acide désoxyribonucléique

L'acide désoxyribonucléique (ADN) est une molécule qui contient toutes les instructions nécessaires à notre croissance, développement, fonctionnement et reproduction. Il est formé de nucléotides, une molécule organique composée d'un acide nucléique (adénine, cytosine, guanine, thymine), d'un sucre (2-désoxyribose) et d'un groupement

phosphate (**Figure 1-6A**). Un brin d'ADN est toujours accompagné d'un brin complémentaire relié par des ponts hydrogènes (ADN double brin ; **Figure 1-6B**) et les deux nucléotides présents à une position précise (un nucléotide pour chaque brin) sont appelés paire de bases. Dans l'ADN double brin, chaque acide nucléique ne peut être relié qu'à un acide nucléique précis. Par exemple, lorsqu'une adénine (A) est présente sur un brin, une thymine (T) est obligatoirement présente sur le brin opposé. Le même phénomène se produit entre les cytosines (C) et les guanines (G). L'ADN double brin est présent sous forme de double hélice dans les cellules (**Figure 1-6C**). Cette conformation confère une plus grande stabilité à l'ADN et aide différentes molécules à se lier à celle-ci.

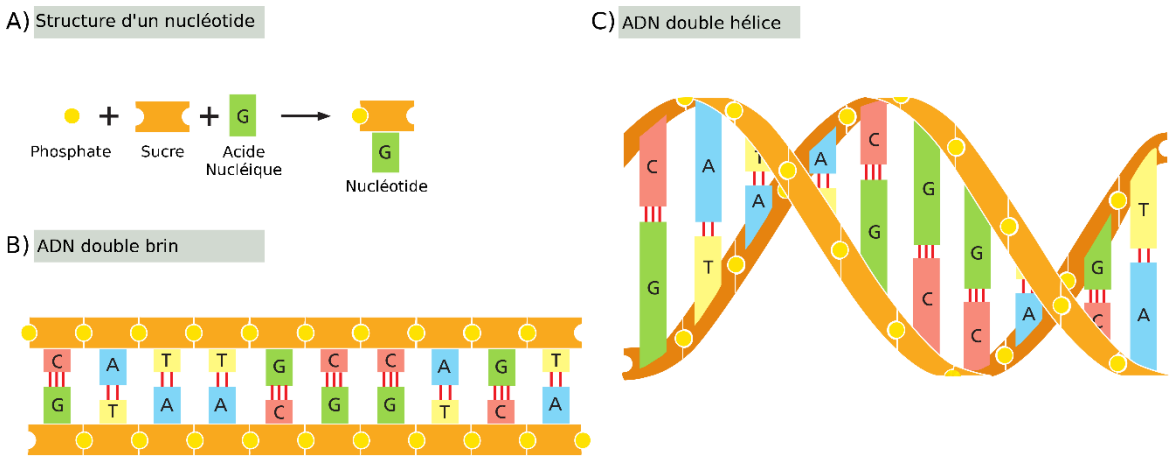

Figure 1-6 : Structure de l'ADN.

Figure adaptée de *Molecular Biology of the Cell 6th edition, 2015*.

1.2.1.2 Chromosome

Un chromosome est une structure compacte et organisée contenant l'ADN (**Figure 1-7**). Les humains ont 22 paires de chromosomes autosomes et deux chromosomes sexuels (X et Y), pour un total de 46. Pour les chromosomes sexuels, les hommes possèdent un chromosome X et un chromosome Y, tandis que les femmes possèdent deux chromosomes X. Chaque parent transmet à son enfant une copie de chaque chromosome non sexuel et un chromosome sexuel. Les femmes transmettent toujours un chromosome X, tandis que les

hommes peuvent transmettre un chromosome X ou un chromosome Y. Les humains possèdent également un génome mitochondrial provenant de la mère.

Les chromosomes sont sous-divisés en plusieurs sections nommées locus (e.g. 4q31.21, 6p21, 15q25). Le nom de chaque locus (loci au pluriel) est représentatif de sa position dans le génome humain. Le premier nombre représente le chromosome où est situé le locus. La lettre suivante correspond au bras sur le chromosome (p est pour le petit bras et q pour le long bras). Les chiffres à droite de la lettre sont la région, la bande et la sous-bande. Par exemple, le locus 4q31.21 se prononce *quatre q trois un point deux un.*

Les chromosomes sont composés d'ADN enroulé autour d'un complexe protéique appelé nucléosome. Ces structures sont elles-mêmes composées de protéines appelées histones. Les chromosomes autosomes sœurs, appelés chromosomes homologues, sont reliés au centre par une structure protéique nommée centromère.

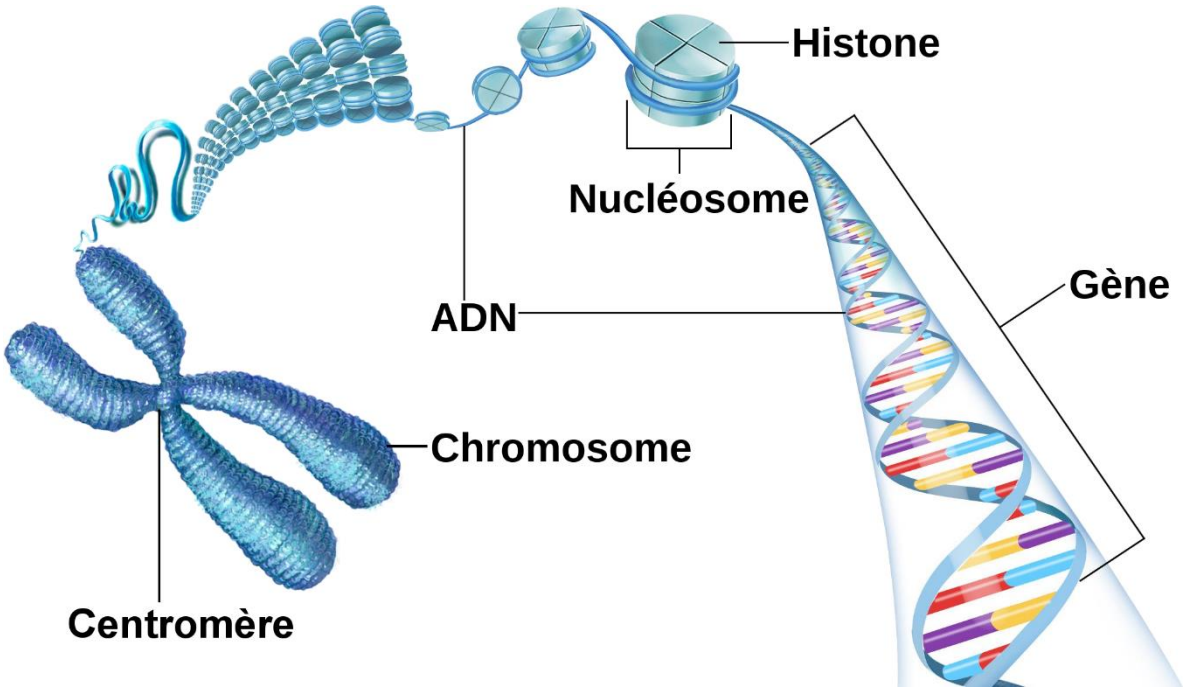

Figure 1-7 : Structure d'un chromosome autosomal.
Figure adaptée du *NIH : National Cancer Institute*.

1.2.1.3 Gène

Un gène est une séquence structurée d'ADN qui sert de manuel à la production d'acide ribonucléique (ARN) messager et de protéines. Il est estimé que le génome humain contient environ 20,000 gènes, mais ce nombre varie selon les définitions employées[271,272]. Les gènes sont principalement composés d'exons et d'introns (**Figure 1-8**). Les exons contiennent les séquences d'ADN qui se retrouvent dans les ARN messager mature. La majorité des exons servira de plan à la production de protéines (exon codant), mais certains sont employés dans la régulation de l'expression et la stabilité des ARN messager (exon non codant). Les introns sont des segments d'ADN non codants situés entre les exons et sont impliqués dans la régulation de l'expression des gènes et dans la stabilité de l'ADN. En plus des exons et des introns, des régions régulatrices se trouvent en amont et en aval des gènes.

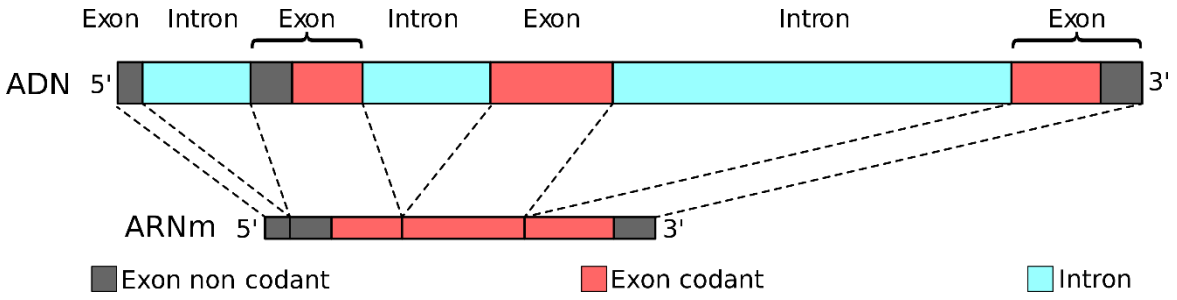

Figure 1-8 : Structure d'un gène.

Figure adaptée de la *Wikipedia foundation*.

Les ARNm et protéines codés par les gènes sont responsables du bon fonctionnement de notre corps en remplissant diverses fonctions (e.g. enzyme, structure, transport). Bien que chaque personne possède les mêmes gènes, des différences existent dans leurs séquences et celles-ci peuvent avoir de graves conséquences sur la santé. Par exemple, des variations dans la séquence codante du gène *CFTR* peuvent causer la fibrose kystique, une maladie qui affecte les épithéliums glandulaires de nombreux organes. Les différents types de variations sont décrits à la section 1.2.1.5.

1.2.1.4 ADN codant

Le génome humain est composé de 3,2 milliards de paires de bases, mais seulement une petite partie (~1%) est de l'ADN codant[273]. La formation d'une protéine à partir de l'ADN s'effectue en plusieurs étapes (**Figure 1-9**). La première étape, nommée transcription, consiste à lire la séquence des exons (exons codants et non codants) pour produire un ARNm. La séquence des ARNm est identique à celle des exons, à l'exception des thymines qui sont remplacées par des uraciles, un acide nucléique spécifique à l'ARN. La deuxième étape traduit la séquence de l'ARNm en acide aminé (traduction). L'identification des acides aminés s'effectue dans les ribosomes, des complexes composés de protéines et d'ARN. Les ribosomes analysent les séquences des ARNm trois nucléotides à la fois. Ces séquences de trois nucléotides consécutifs sont nommées codon et chaque codon code pour seulement un acide aminé. Chez l'humain, les 64 codons possibles codent pour un total de 20 acides aminés et trois codons-stop (**Tableau 1-4**). Outre le tryptophane et la méthionine, tous les acides aminés sont associés à plusieurs codons. Par exemple, la phénylalanine est encodée par les codons TTT et TTC. Cette redondance dans le code génétique est appelée dégénérescence du code génétique[274]. L'ordre dans lequel le ribosome lit la séquence de codons se nomme cadre de lecture. Certaines mutations peuvent avoir un impact sur le cadre de lecture. Une fois la traduction de l'ARNm complétée, la chaîne d'acides aminés subit différentes étapes de modifications avant de devenir une protéine.

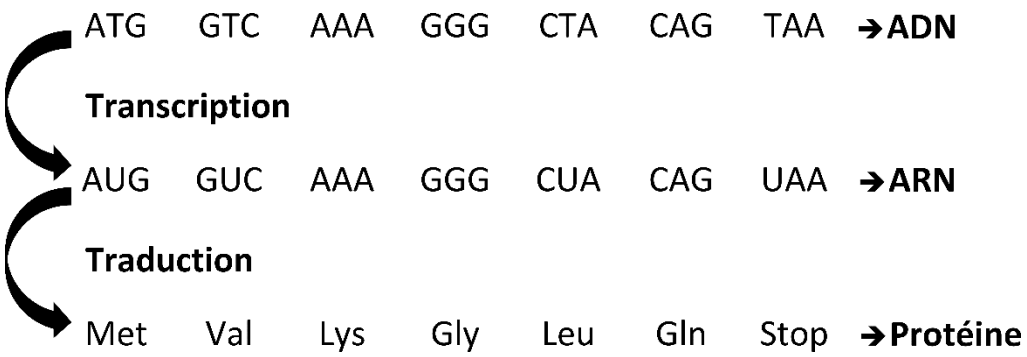

Figure 1-9 : Transcription et traduction.

Tableau 1-4 : Code de l'ADN.

1ère base	2e base								3e base
	T		**C**		**A**		**G**		
T	TTT	F - Phe Phénylalanine	TCT	S - Ser Sérine	TAT	Y - Tyr Tyrosine	TGT	C - Cys Cystéine	**T**
	TTC	F - Phe Phénylalanine	TCC	S - Ser Sérine	TAC	Y - Tyr Tyrosine	TGC	C - Cys Cystéine	**C**
	TTA	L - Leu Leucine	TCA	S - Ser Sérine	TAA	Stop	TGA	Stop	**A**
	TTG	L - Leu Leucine	TCG	S - Ser Sérine	TAG	Stop	TGG	W - Trp Tryptophane	**G**
C	CTT	L - Leu Leucine	CCT	P - Pro Proline	CAT	H - His Histidine	CGT	R - Arg Arginine	**T**
	CTC	L - Leu Leucine	CCC	P - Pro Proline	CAC	H - His Histidine	CGC	R - Arg Arginine	**C**
	CTA	L - Leu Leucine	CCA	P - Pro Proline	CAA	Q - Gln Glutamine	CGA	R - Arg Arginine	**A**
	CTG	L - Leu Leucine	CCG	P - Pro Proline	CAG	Q - Gln Glutamine	CGG	R - Arg Arginine	**G**
A	ATT	I - Ile Isoleucine	ACT	T - Thr Thréonine	AAT	N - Asn Asparagine	AGT	S - Ser Sérine	**T**
	ATC	I - Ile Isoleucine	ACC	T - Thr Thréonine	AAC	N - Asn Asparagine	AGC	S - Ser Sérine	**C**
	ATA	I - Ile Isoleucine	ACA	T - Thr Thréonine	AAA	K - Lys Lysine	AGA	R - Arg Arginine	**A**
	ATG	M - Met Méthionine Initiation	ACG	T - Thr Thréonine	AAG	K - Lys Lysine	AGG	R - Arg Arginine	**G**
G	GTT	V - Val Valine	GCT	A - Ala Alanine	GAT	D - Asp Acide aspartique	GGT	G - Gly Glycine	**T**
	GTC	V - Val Valine	GCC	A - Ala Alanine	GAC	D - Asp Acide aspartique	GGC	G - Gly Glycine	**C**
	GTA	V - Val Valine	GCA	A - Ala Alanine	GAA	E - Glu Acide glutamique	GGA	G - Gly Glycine	**A**
	GTG	V - Val Valine	GCG	A - Ala Alanine	GAG	E - Glu Acide glutamique	GGG	G - Gly Glycine	**G**

1.2.1.5 Variation génétique

En moyenne, le code génétique entre chaque humain est identique à 99.5%[275]. Les variations génétiques sont responsables des différences observées entre individus d'une même espèce. Selon la version 147 de dbSNP (base de données des variations connues), plus de 150 millions de variations différentes ont été répertoriées chez l'humain. Plusieurs

types de variations existent dans le génome humain et celles-ci peuvent avoir des conséquences très diverses selon leur localisation.

Polymorphisme nucléotidique simple (SNP)

Un polymorphisme nucléotidique simple (SNP) est une variation d'une seule paire de base du génome observée entre individus d'une même espèce. Ils sont le type de variation le plus fréquent dans le génome humain[276,277]. Les SNP peuvent avoir différents impacts selon leur localisation (**Figure 1-10**) :

- Non codant - Intergénique : Situé dans une région non codante entre deux gènes. Peut affecter l'expression de gènes.
- Non codant - Exon : Présent dans un exon non codant d'un gène. La protéine demeure inchangée, mais peut affecter l'expression de gènes.
- Codant - Synonyme ou silencieux (*synonymous*) : Situé dans un exon codant. L'acide aminé ne change pas et la protéine demeure inchangée.
- Codant - Faux sens (*missense*) : Situé dans un exon codant. L'acide aminé change et la protéine est modifiée.
- Codant - Non-sens (*nonsense*) : Situé dans un exon codant. Un codon-stop apparaît. La protéine est tronquée et généralement non fonctionnelle.

Séquence de référence

ATG	GTC	AAA	GGG	CTA	CAG	TAA
Met	Val	Lys	Gly	Leu	Gln	Stop

Synonyme

ATG	GTC	AAG	GGG	CTA	CAA	TAA
Met	Val	Lys	Gly	Leu	Gln	Stop

Faux sens

ATG	GTC	AAT	GGG	CTA	CAT	TAA
Met	Val	Asn	Gly	Leu	His	Stop

Non-sens

ATG	GTC	TAA
Met	Val	Stop

Figure 1-10 : Types de polymorphisme nucléotidique simple (SNP) codant.

Insertion et délétion (Indel)

Les insertions et les délétions (indel) se caractérisent par l'ajout ou le retrait d'une ou de plusieurs bases (**Figure 1-11**). Si un indel se produit dans une séquence codante, cela peut avoir de graves conséquences pour la protéine. Un indel dont la taille n'est pas un multiple de trois produit une mutation du cadre de lecture (*frameshift mutation*). Lorsque ce type de mutation est présent, la séquence des codons et la chaîne d'acides aminés est modifiée. Cela a comme résultat une protéine avec une capacité fonctionnelle différente ou une protéine non fonctionnelle. Lorsque la taille d'un indel est un multiple de trois, la séquence d'acides aminés est soit tronquée ou allongée.

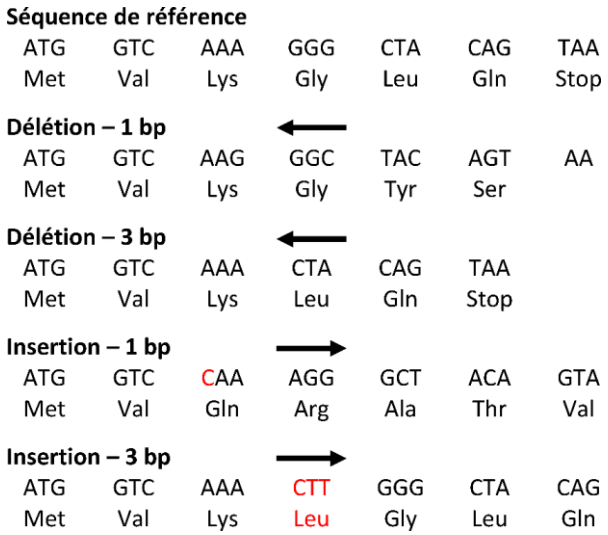

Figure 1-11 : Exemples d'insertions et de délétions.

Variabilité du nombre de copies

Une variabilité du nombre de copies est l'absence ou la présence d'une ou plusieurs copies d'une région chromosomique de 50 paires de bases ou plus[278,279]. Ce type de variation, mieux connue sous le nom de *copy number variation* (CNV), peut produire de nombreux scénarios ; absence d'un gène, une seule copie d'un gène, plusieurs copies d'un gène, gène non fonctionnel, fusion de gènes.

1.2.1.6 Technologies et approches en génomique

Avec l'avancement de nos connaissances en génétique, de nouvelles technologies et approches furent développées afin d'interroger le génome humain. Ces outils ont permis de faire d'important progrès pour la génétique des maladies humaines. Les paragraphes suivants présenteront quelques-unes des techniques les plus couramment employées.

1.2.1.6.1 Puce à ADN (microarray)

Une puce à ADN, mieux connu sous le nom de *DNA microarray*, est un ensemble de molécules d'ADN fixé de façon ordonnée sur une surface solide[280]. Ces séquences, qui sont connues au départ, permettent d'analyser des sites précis sur le génome humain. Au départ, cette technologie était employée pour quantifier l'expression de milliers de gènes. Aujourd'hui, les puces à ADN permettent également d'analyser une vaste gamme de phénomènes biologiques dont les SNPs, la méthylation, les gènes de fusion et la détection de pathogènes[281]. Cette technologie peut tout autant être employée chez l'humain que chez la souris, le maïs, et d'autres espèces (https://www.illumina.com/). Les puces à ADN sont régulièrement employées dans les études pangénomiques en raison de leur rapport coût/quantité de données[282]. Par contre, cela risque de changer au cours des prochaines années avec la diminution fulgurante des coûts du séquençage[283].

1.2.1.6.2 Séquençage de l'ADN

Le séquençage de l'ADN consiste à déterminer la séquence d'acides nucléiques d'un brin d'ADN. La technique la plus connue et la plus employée est la méthode de Sanger[284]. Le séquençage est employé de différente façon, mais principalement pour séquencer des régions précises du génome humain (*targeted sequencing*), tous les exons du génome (*whole exome sequencing*) ou le génome dans son entièreté (*whole genome sequencing*). Contrairement aux puces à ADN, le séquençage de l'ADN détermine tous les génotypes d'une séquence et non seulement ceux interrogés sur une puce. Par contre, cette technologie est plus dispendieuse et la quantité de données générées complique l'analyse des résultats[283].

1.2.1.6.3 Genome-wide association study (GWAS)

Les criblages génomiques par association, mieux connus sous le nom de *genome-wide association study* (GWAS), sont un type d'étude génétique qui consiste à vérifier l'association entre plusieurs centaines de milliers de polymorphismes et un phénotype précis[285] (**Figure 1-12**). Ce genre d'étude nécessite des cohortes composées d'individus avec (cas) ou sans (témoins) un phénotype précis (e.g. MPOC, cancer) ou avec de l'information sur une caractéristique clinique continue (e.g. VEMS, grandeur). Peu importe le phénotype étudié, les cohortes doivent être composées de centaines voir de milliers d'individus avec les informations cliniques et génétiques pour tous les sujets. Suite aux analyses statistiques, des variations génétiques associées avec le phénotype sont identifiées.

Bien que les criblages génomiques par association soient une technique très puissante, de nombreuses limitations compliquent l'interprétation des résultats. Un des défis techniques des GWAS est le recrutement d'un nombre important de sujets. Les chercheurs ont alors recours à des sujets provenant de plusieurs centres qui peuvent être situés dans différents pays. Cela peut poser problème puisque l'ethnicité des sujets peut varier entre les centres, et les résultats de GWAS sont connus pour varier entre les groupes ethniques[286]. Une autre limitation est la couverture génétique. La majorité des GWAS effectués jusqu'à présent ont employé des puces à ADN. Bien que celles-ci permettent d'identifier les génotypes de centaines de milliers ou de millions de variations génétiques, un nombre important de variations est ignoré et particulièrement les variations rares[287,288]. Finalement, les résultats obtenus lors de ces études sont difficiles à interpréter. La majorité des variations identifiées par GWAS n'expliquent que faiblement le phénotype étudié[288] et sont localisées dans des régions non codantes (intron, intergénique)[289–291]. Cette dernière observation complique grandement l'identification du ou des gènes causaux. Des études fonctionnelles supplémentaires sont donc essentielles pour identifier les variations et gènes causaux.

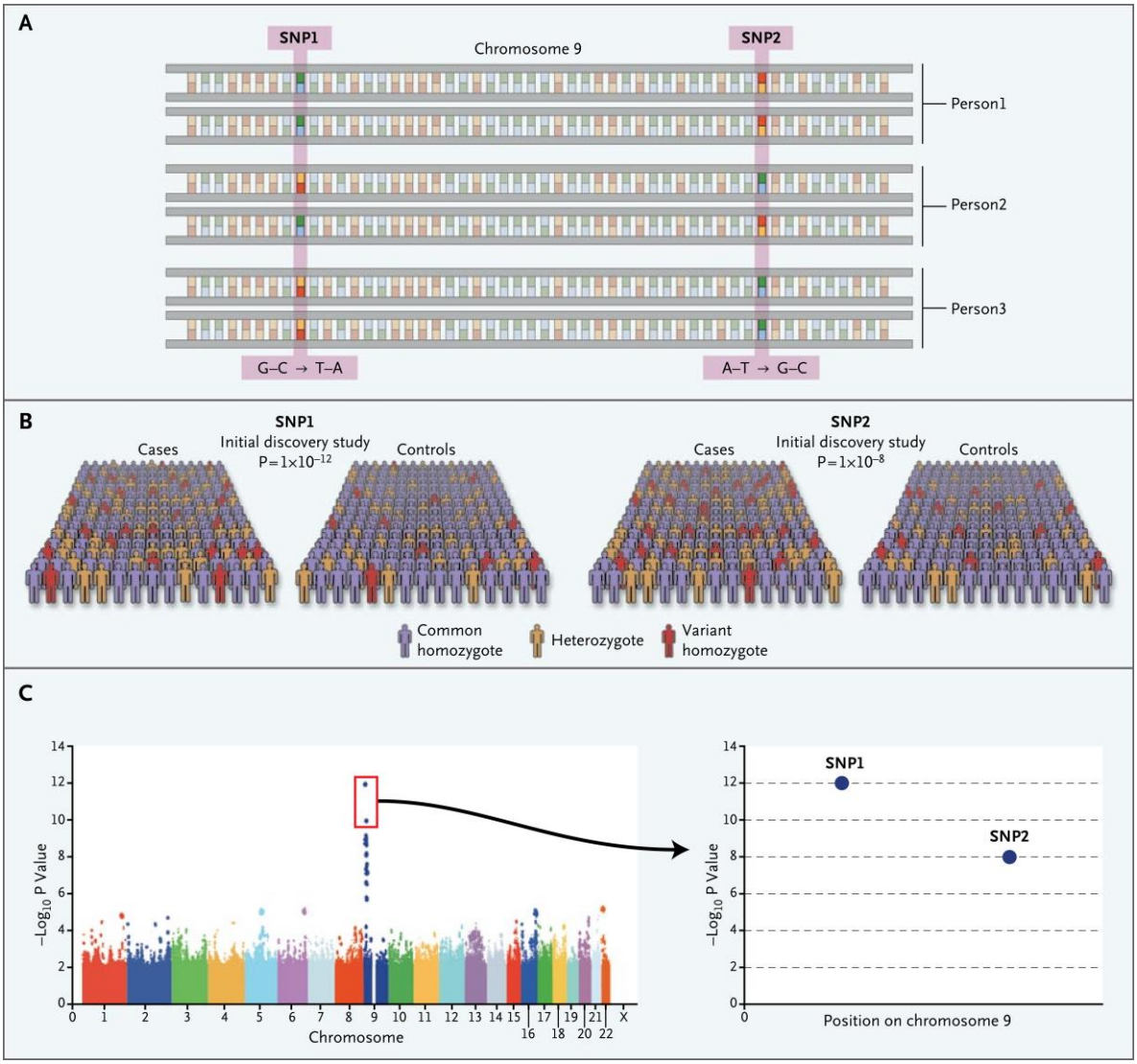

Figure 1-12 : Exemple de criblage génomique par association (GWAS).

Figure de T. Manolio, *Genomewide association studies and assessment of the risk of disease*, *N Engl J Med* 2010 Jul 8;363(2):166-76[285]. A) Un locus du chromosome 9 contenant deux variations génétiques (SNP1 et SNP2). B) Distribution des génotypes des SNP1 et SNP2 dans une cohorte de sujets cas et une de sujets témoins. C) Exemple de graphique de Manhattan (*Manhattan plot*). L'axe des X est la position des SNP sur le génome humain. L'axe des Y est la valeur P transformée à la base -log10. Chaque point est le résultat d'un test statistique entre un SNP et le phénotype étudié. Le graphique de droite est un zoom des deux variations les plus significatives dans cet exemple.

1.2.1.6.4 Expression Quantitative Trait Loci (eQTL)

Les *expression Quantitative Trait Loci* (eQTL) sont des variations génétiques influençant l'expression des gènes (**Figure 1-13**). Pour détecter un eQTL, il est nécessaire d'avoir chez le plus grand nombre d'individus possible, l'expression des gènes et les génotypes de variations génétiques. Il est important de préciser que pour chaque sujet, les deux types de données doivent être disponibles. Les tests statistiques effectués permettent de déterminer si les différents génotypes d'une variation expliquent les changements observés dans les niveaux d'expression d'un gène.

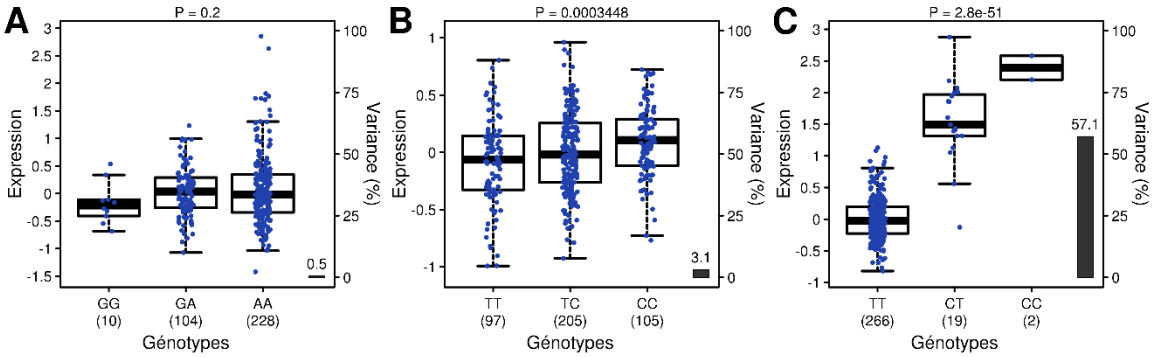

Figure 1-13 : Exemple d'expression Quantitative Trait Loci.

L'axe des x représente les génotypes de la variation génétique. L'axe des y de gauche représente l'expression du gène. L'axe des y de droite représente la variance de l'expression expliquée par la variation génétique. A) Aucun eQTL. B) Un eQTL significatif, mais avec un effet relativement faible. C) Un eQTL significatif expliquant un grand pourcentage de la variance.

Les eQTL sont régulièrement employés pour identifier les gènes causaux dans les régions chromosomiques identifiées par les GWAS[292]. Les criblages génomiques permettent d'identifier des variations associées à un phénotype précis, mais dans la majorité des cas ne peuvent déterminer le ou les gènes causaux[291]. De plus, les variations détectées par ce type d'étude sont souvent localisées dans des régions non codantes. En déterminant quel gène est influencé par une variation génétique précédemment associée, il

devient alors possible de déterminer les mécanismes moléculaires précis. L'article publié par Kapoor *et al*.[293] en 2014 est un exemple parfait de l'utilité des eQTL. Ils ont démontré que le variant intergénique rs7539120 influence les fonctions cardiaques en régulant à la hausse l'expression du gène *NOS1AP*[293], un gène précédemment associé à ce phénotype.

Les eQTL se subdivisent en deux catégories dépendamment de la distance entre la variation génétique et le gène associé (**Figure 1-14**). Lorsque le SNP et le gène sont situés à proximité (< 1 Mb), nous parlons d'un eQTL local (*local*-eQTL) ou d'un *cis*-eQTL[294]. Nous parlons d'eQTL distant (*distant*-eQTL) ou de *trans*-eQTL quand la distance est supérieure à 1 Mb ou que la variation et le gène se situent sur des chromosomes différents. Par contre, les termes employés peuvent varier d'une étude à l'autre[295]. La distinction entre les eQTL local et distant est très importante, car les mécanismes impliqués ne sont pas les mêmes. Pour les local-eQTL, les mécanismes possibles sont nombreux et variés[295] (e.g. liaison du facteur de transcription, méthylation de l'ADN, accessibilité de la chromatine, épissage alternatif, dégradation de l'ARNm). L'exemple présenté dans le paragraphe précédent est un eQTL local. Le variant identifié, rs7539120, est situé dans un amplificateur (*enhancer*) situé à 20 kb du gène *NOS1AP*. Pour ce qui est des eQTL distants, un mécanisme intermédiaire est responsable de la différence observée[295]. Une variation génétique présente sur le chromosome 1 pourrait influencer l'expression d'un ou de plusieurs gènes situés sur des chromosomes différents en influençant l'expression d'un facteur de transcription situé à proximité. Par exemple, Lee *et al*.[296] ont remarqué dans les cellules dendritiques que le polymorphisme rs12805435, localisé sur le chromosome 11, est un eQTL local pour le gène *IRF7* et un eQTL distant pour sept gènes additionnels[296] (*NMI*, *IFNA4*, *IFNA10*, *IFNA13*, *IFNA21*, *IFNA17* et *IFNA5*) situés sur les chromosomes 2 et 9. Ce eQTL distant se produit car *IRF7* est un facteur de transcription pour les autres gènes.

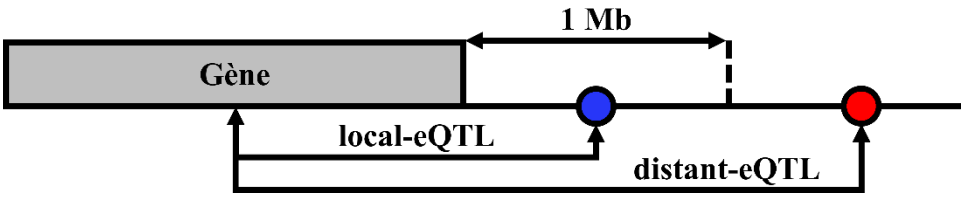

Figure 1-14 : eQTL local et distant.

Le rectangle gris représente la position du gène. Le cercle bleu est un eQTL local pour le gène et le cercle rouge est un eQTL distant.

Tout comme les GWAS, les eQTL ont quelques limitations. Bien que les eQTL peuvent être partagés par plusieurs tissus[297,298], ceux-ci peuvent également être tissus spécifiques[299,300]. Une association détectée dans un tissu (e.g. sang) peut donc être absente dans un autre tissu (e.g. poumon). Cela complique grandement les études sur les maladies dont les tissus impliqués sont difficilement accessibles (e.g. poumon, cœur, cerveau). Une autre limitation est que les niveaux d'expression ne corrèlent pas nécessairement avec les niveaux de protéines[301]. Une variation qui affecte les niveaux d'expression d'ARNm peut n'avoir aucun impact sur la quantité de protéine produite. Finalement, les eQTL ne donnent aucune information sur les mécanismes de régulation impliqués. Des études supplémentaires sont donc nécessaires pour comprendre comment la variation génétique influe sur les niveaux d'ARNm.

1.2.2 Génétique de la MPOC

Tel que mentionné à la section 1.1, pas tous les fumeurs développeront une obstruction respiratoire. De plus, la maladie se manifeste différemment d'un patient à l'autre[302] (e.g. présence ou absence de bronchite chronique et d'emphysème, nombre de comorbidités). Ces observations combinées à des études familiales ont démontré que la génétique est impliquée dans le développement de la maladie[3,303]. Malgré l'association d'un grand nombre de variations génétiques et de gènes avec la MPOC, seul le gène *SERPINA1* est connu pour causer la maladie sans exposition aux produits du tabac. Les sections suivantes feront un résumé de l'état actuel de nos connaissances sur la génétique de la maladie.

1.2.2.1 Déficit en alpha-1 antitrypsine

Le déficit en alpha-1 antitrypsine, identifié pour la première fois en 1963 par Carl-Bertil Laurell et Sten Eriksson[304], est causé par des mutations dans la séquence codante du gène *SERPINA1*. Ces mutations ont pour effet un arrêt prématuré de la transcription du gène (mutation non-sens) ou la production d'une protéine non fonctionnelle (mutation faux sens). L'AAT est une antiprotéase principalement produite par le foie, mais également par les neutrophiles, les monocytes et les cellules épithéliales du poumon et du système digestif[305]. Une fonction de cette protéine est de protéger les poumons des effets destructeurs de l'élastase neutrophile en l'inhibant[306]. Tel que mentionné à la section 1.1.5.2, les personnes souffrant d'un déficit peuvent développer de l'emphysème au cours de leur vie[307]. Ces individus sont également à risque de développer d'autres pathologies pulmonaires (e.g. asthme[308], bronchectasie[309]) et de souffrir de maladies du foie[242,243] telles que la cirrhose hépatique.

Pour le gène *SERPINA1*, la littérature scientifique emploie le terme variant pour décrire le comportement de la protéine. Au départ, les variants du gène étaient nommés selon leur vitesse de migration lors d'une électrophorèse sur gel : M – *medium* (médium), S – *slow* (lent), F – *fast* (rapide), Z – *very slow* (très lent)[312]. Avec la venue du génotypage et du séquençage, les variations génétiques sont analysées afin de déterminer les variants. De plus, les variants détectés sont nommés Pi (*protease inhibitor*), basé sur l'ancien nom du locus où est situé le gène [305,312]. Actuellement, les variants sont séparés en trois catégories selon leur impact sur les niveaux sanguins de la protéine : Normal, Déficient, Nul. Les variants du type normal sont caractérisés par des niveaux normaux d'AAT et n'augmentent pas le risque de développer de l'emphysème ou autres conditions associées. Les variants Pi M1 à M4 sont les variants normaux les plus communs et sont présents dans la majorité de la population[313–315]. Les variants du type déficient sont associés à des niveaux moindres de protéine dans le sérum sanguin et à des risques accrus de développer de l'emphysème. Les deux variants déficients les plus communs sont Pi Z et S[305,312]. Les variants nuls, aussi appelées Q0, sont caractérisés par l'absence de niveau détectable de la protéine et sont également associées à un risque accru de développer de l'emphysème[316]. Bien que le gène soit étudié depuis plus de 50 ans, de nouvelles mutations/variants pathogéniques sont découvertes continuellement[316–318].

Tel que mentionné dans le paragraphe précédent, les variants Pi Z et S sont les variants pathogéniques les plus communs[220]. Le variant Pi Z est composé d'une lysine à la position 342 de la protéine (Glu342Lys; rs28929474). Ce variant produit une protéine mal repliée qui s'accumule dans le réticulum endoplasmique des cellules[319]. Plusieurs études ont estimé l'âge de ce variant, mais celui-ci varie grandement entre les études (2,000 à 6,000 ans)[320–323]. Le deuxième variant déficient le plus courant, Pi S, est causé par le changement d'un acide glutamique en une valine à la position 264 (Glu264Val; rs17580). Comparativement au variant Z, le variant S a un impact moindre sur la concentration de la protéine dans le sang. D'autres variants déficients sont connus pour causer un déficit (e.g. F, I, P_{Lowell}), mais ils sont beaucoup plus rares[176].

Selon de récentes études épidémiologiques, la prévalence du déficit varie grandement entre les régions du monde[324]. Par exemple, l'Asie est très peu touchée par cette maladie puisque la fréquence des variants Pi Z et S est extrêmement faible chez les asiatiques. À l'opposé, les régions du monde principalement colonisées par les pays du nord de l'Europe (e.g. Danemark, Norvège, Lettonie) sont plus fortement affectées par le déficit.

1.2.2.2 Études par gènes candidats

Lors des dernières décennies, plusieurs études ont été réalisées à l'aide de l'approche par gènes candidats[8]. Ce genre d'étude cible les gènes de voies biologiques soupçonnées de jouer un rôle dans le phénotype d'intérêt. Les principales voies biologiques étudiées pour la MPOC ont été celles de l'inflammation, du système immunitaire, du stress oxydant, du métabolisme des xénobiotiques et de l'équilibre protéase-antiprotéase[325]. Les études par gènes candidats ont permis, avant la venue des GWAS, d'identifier plusieurs gènes et régions chromosomiques possiblement associés à la MPOC (e.g. *GSTM1*, *TGFB1*, *TNF*)[8]. Par contre, les résultats obtenus ont été difficile à répliquer d'une étude à l'autre. Un bon exemple de cette situation est le gène *TNF* (*Tumor necrosis factor alpha*)[8]. Plus d'une vingtaine d'études ont établi une association positive entre des variations génétiques de ce gène et la MPOC. Par contre, un nombre équivalent d'études n'a observé aucun lien entre le gène et la maladie. Actuellement, les études par gènes candidats sont régulièrement employées pour confirmer et raffiner les résultats obtenus par les GWAS[326].

1.2.2.3 Genome-wide association study

Lors des dernières années, plusieurs études pangénomiques effectuées sur la MPOC ont permis d'associer des variations génétiques et des régions chromosomiques avec la maladie (**Tableau 1-5**). Ces études réalisées avec des milliers d'individus et différents groupes ethniques ont permis de relier 24 loci avec la MPOC[90,327–332] et plus d'une centaine avec des phénotypes associés (e.g. bronchite chronique, cachexie, diminution des fonctions respiratoires, emphysème, fonction pulmonaire, réactivité des voies respiratoires)[93,333–360]. Parmi les régions chromosomiques identifiées, certaines l'ont été dans plusieurs d'études et parfois avec des phénotypes différents. Quatre de ces régions (4q22, 4q31, 6p21.32, 15q25) sont décrites dans les paragraphes suivants.

Tableau 1-5 : GWAS effectués sur la MPOC.

Références	**Nombre de sujets (cas vs témoins)**	**Nombre de régions identifiées**	**Nombre de régions connues**
Pillai *et al*. PLoS Genet. 2009.[90]	823 / 810	2	2
Cho *et al*. Nat Genet. 2010.[329]	2940 / 1380	3	3
Cho *et al*. Hum Mol Genet. 2011.[330]	3499 / 1922	4	4
Cho *et al*. Lancet Respir Med. 2014.[328]	6633 / 5704	6	7
Chen *et al*. Ann Am Thorac Soc. 2015.[327]	157 / 1031	2	9
Hobbs *et al*. Am J Respir Crit Care Med. 2016.[331]	6161 / 6004	3	10
Hobbs *et al*. Nat Genet. 2017.[332]	15,256 / 47,936	21	24

1.2.2.3.1 4q22 – *FAM13A*

La première association entre ce locus et la MPOC fut réalisée en 2010 par Cho *et al*[329]. Depuis, quatre GWAS supplémentaires ont confirmé cette association [328,330,332,344,361] et le locus 4q22 fut également associé aux fonctions respiratoires[342,346,352,353,360] et à la bronchite chronique[344]. *FAM13A* (*Family With Sequence Similarity 13 Member A*) est le gène candidat dans cette région. Malgré de nombreuses associations entre ce gène et la MPOC, son rôle dans le développement de la maladie n'est toujours pas compris avec certitude. La théorie actuelle suggère que le gène serait impliqué dans la voie de signalisation des Rho GTPase, des régulateurs clés du cytosquelette impliqués dans plusieurs maladies pulmonaires[362]. Chez les fumeurs susceptibles, le gène *FAM13A* serait surexprimé dans les poumons[337,363] et ceci

activerait la dégradation de la bêta-caténine, une protéine impliquée dans la réparation cellulaire des voies respiratoires et alvéoles pulmonaires[364]. Une présence réduite de cette protéine chez les individus fumeurs serait donc associée à une diminution de la réparation cellulaire dans les poumons.

1.2.2.3.2 4q31 – *HHIP*

La première association entre le locus 4q31 et la MPOC fut réalisée en 2009 par Pillai *et al*[90]. Depuis, ce locus a été associé à la MPOC dans quatre autres études pangénomiques[328–330,332]. La région 4q31 a également été identifiée dans quelques GWAS sur les fonctions respiratoires[334,337,340,352,360] et l'emphysème[357,358]. Le gène candidat de ce locus, *HHIP* (*hedgehog interacting protein*), code une protéine fortement conservée. Une étude parue en 2012 a démontré que ce gène est sous-exprimé dans le poumon de patients MPOC[365]. Ce gène est impliqué dans les voies biologiques de la matrice extracellulaire et croissance cellulaire[366], deux voies associées à la MPOC[367]. Une diminution de l'expression pourrait donc participer au développement de la MPOC en dérégulant ces deux voies biologiques.

1.2.2.3.3 6p21.32 – *AGER*

Contrairement aux deux régions chromosomiques précédentes, le locus 6p21.32 a principalement été associé aux fonctions respiratoires (VEMS, VEMS/CVF) [336,339,340,346,352,353,360]. En 2014 et 2015, deux GWAS ont établi un lien entre cette région et l'emphysème[341,357]. Finalement en 2017, une étude pangénomique sur la MPOC a établi un lien entre ce locus et la maladie[332]. La protéine produite par le gène candidat, *AGER* (*advanced glycosylation end-product specific receptor*), est impliquée dans l'inflammation, l'homéostasie et le développement[368]. Une étude parue en 2011 a démontré que les niveaux de *RAGE* étaient supérieurs dans les voies respiratoires et poumons de sujets MPOC[369]. *AGER*/*RAGE* participent probablement au développement de l'obstruction respiratoire en régulant à la hausse l'inflammation et l'apoptose[370].

1.2.2.3.4 15q25 – *CHRNA3, CHRNA5, CHRNB4, IREB2*

Parmi tous les loci identifiés par GWAS, la région chromosomique 15q25 est la plus constante. Celle-ci a été identifiée dans six GWAS sur la MPOC[90,328–332], un sur les

fonctions respiratoires[93] et deux sur l'emphysème[357,358]. De plus, cette région est également associée à la consommation tabagique[371–374], le principal facteur de risque de la MPOC, et au cancer du poumon[91,92,375–379], une comorbidité importante. Cette région comporte de nombreux gènes susceptibles d'être causaux (*CHRNA3*, *CHRNA5*, *CHRNB4*, *IREB2*, *PSMA4*). *CHRNA3* (*cholinergic receptor nicotinic alpha 3*), *CHRNA5* (*cholinergic receptor nicotinic alpha 5*) et *CHRNB4* (*cholinergic receptor nicotinic beta 4 subunit*) sont tous des sous-unités du récepteur nicotinique de l'acétylcholine (*nicotinic acetylcholine receptors*) qui est impliqué dans la dépendance au tabac[380,381]. Ces gènes sont soupçonnés de participer au développement de la MPOC en influençant la consommation tabagique. *IREB2* (*iron responsive element binding protein 2*) est impliqué dans le métabolisme du fer[382]. Une étude parue en 2009 a démontré que les niveaux de mRNA et de protéine de ce gène sont supérieurs chez les patients MPOC comparativement à des patients contrôles[383]. Une augmentation des niveaux de fer dans les poumons pourrait contribuer au stress oxydant important observé chez ces individus[384]. Ces résultats pointent vers deux mécanismes indépendants pouvant contribuer au développement de la MPOC dans ce locus.

1.2.2.3.5 Fonction respiratoire

Parmi les GWAS conduits sur des phénotypes associés à la MPOC, ceux des fonctions respiratoires ont été les plus nombreux et les plus prolifiques. Suite à la publication de Wain *et al*. en 2017[360], le nombre de régions chromosomiques associées aux fonctions respiratoires (VEMS, CVF, VEMS/CVF) est passé de 54 à 97 (**Figure 1-15**). Les signaux pangénomiques connus expliquent respectivement 9.6%, 6.4% et 14.3% de l'héritabilité de ces phénotypes, qui est estimée à 40% pour chacun[385,386].

Tel que mentionné précédemment, la MPOC et les fonctions respiratoires partagent de nombreuses régions de susceptibilité et cette étude ne fait pas exception. Outre l'identification de régions précédemment associées à la MPOC (*FAM13A*, *AGER*, *GSTCD*), les auteurs ont également démontré que les variations associées aux fonctions respiratoires étaient pour la plupart également associées à la MPOC. Cette étude démontre encore une fois le lien important entre la génétique de ces deux phénotypes.

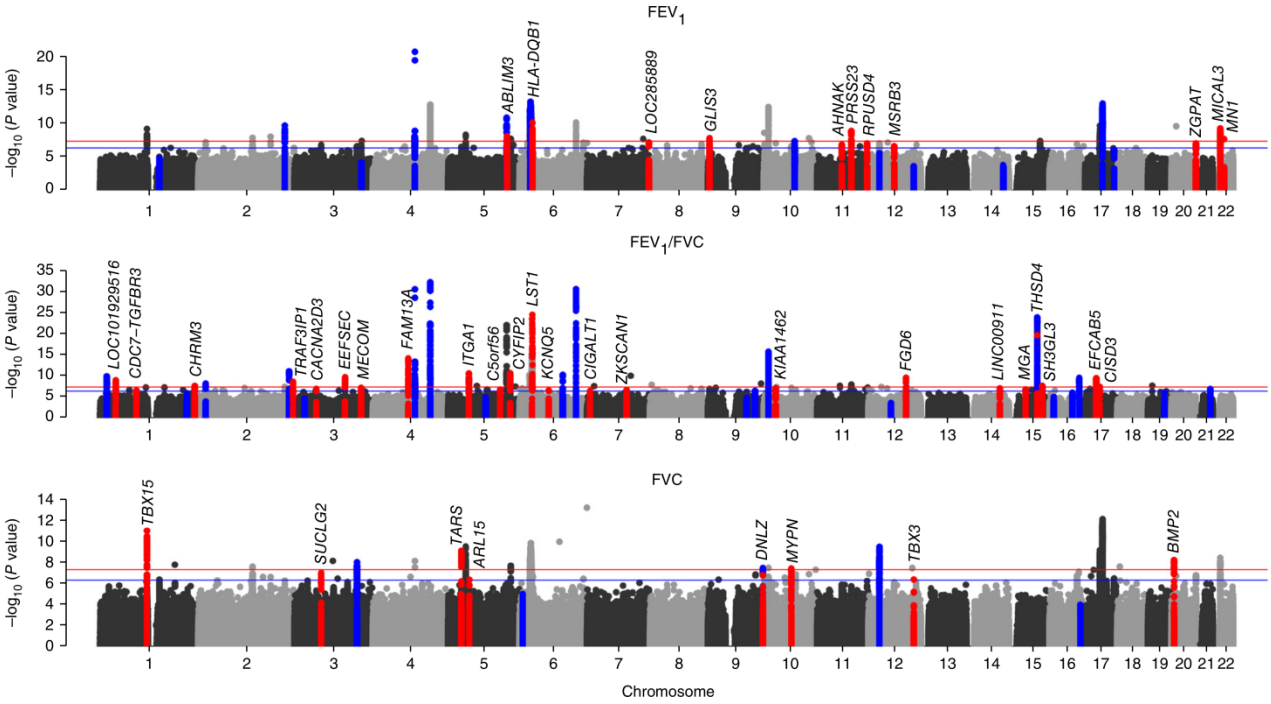

Figure 1-15 : Graphique de Manhattan de Wain *et al.*, *Nat Gen*, 2017.

Figure de L. Wain, *Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets*, *Nat Genet* 2017 Mar;49(3):416-425[285]. Graphique de Manhattan représentant les résultats significatifs de l'étude pour le VEMS, le ratio VEMS/CFV et la CVF. Les nouveaux loci sont en rouge et ceux précédemment identifiés en bleu.

1.2.2.4 expression Quantitative Trait Loci

Dans la littérature scientifique, peu d'études se sont attardées à identifier des eQTL en relation avec la MPOC[363,387–390]. La majorité d'entre elles ont employé les données du consortium d'eQTL pulmonaire pour réaliser ces analyses. Dans les paragraphes suivants, le consortium d'eQTL pulmonaire ainsi que deux autres études seront décrits en profondeur.

1.2.2.4.1 Consortium d'eQTL pulmonaire

Le consortium d'eQTL pulmonaire (*Lung eQTL consortium*) est composé de chercheurs provenant de trois universités (Université Laval, Canada; Université de la Colombie-Britannique, Canada; Université de Groningen, Pays-Bas) et d'une compagnie pharmaceutique (Merck). Chez près de 1,500 individus ayant subi une résection pulmonaire, du tissu pulmonaire non tumoral et des échantillons sanguins ont été prélevés. L'ARNm des échantillons pulmonaires a été hybridé sur des *microarray* personnalisés de la compagnie Affymetrix. L'ADN provenant soit de tissu pulmonaire ou d'échantillon sanguin a été testé sur des puces à ADN Human1M-Duo de la compagnie Illumina. Suivant les étapes de contrôle de qualité, 1,111 sujets sont disponibles pour effectuer des analyses d'eQTL. Les données générées par ce consortium ont été employées dans un grand nombre d'études sur plusieurs phénotypes pulmonaires, dont l'asthme[389,391], le cancer du poumon[389,392], la fibrose kystique[389,393], les fonctions respiratoires[360,389,394] et la MPOC[332,363,390,395,396].

1.2.2.4.2 Génétique de l'expression des gènes dans les expectorations de patients MPOC

Cette étude de Qiu *et al*. parue en 2011 a identifié des eQTLs dans les expectorations induites de 131 sujets MPOC[387]. Vu le petit nombre de sujets pour ce type d'étude, les eQTL calculés se limitaient à une distance de plus ou moins 50 kilobases du site d'initiation de la transcription des gènes. Le résultat principal de cet article est que parmi les 3,309 SNP impliqués dans les eQTL, deux variations sont également associées à la MPOC dans une cohorte de 2,940 cas et 1,380 témoins. Ces deux SNP, rs1051730 et rs2656069, sont des eQTL pour deux gènes précédemment associés à la MPOC, soit *CHRNA5* et *IREB2* respectivement. Malgré le petit nombre d'échantillons pour des

analyses d'eQTL, cette étude a démontré que les eQTL ont la capacité d'identifier de possibles gènes causaux.

1.2.2.4.3 Contrôle génétique de l'expression des gènes dans les loci de la MPOC

En 2014, Castaldi *et al*. ont publié une étude qui s'est attardée à identifier des eQTL dans les expectorations induites et dans le sang de patients MPOC[388]. Cet article, qui est la suite de celui de Qiu *et al*.[387], a effectué des analyses d'intégration entre des données de GWAS et d'eQTL. Le résultat principal de ce projet est l'identification de 19 régions significatives à la fois dans le GWAS et les eQTL sanguins et/ou d'expectorations. Parmi celles-ci, les loci 4q22, 4q31 et 15q25, décrits aux sections 1.2.2.3.1/2/4, ont été identifiés. Autre résultat intéressant, une forte proportion des variations génétiques se situait dans les sites d'initiation de la transcription des gènes. Bien que cette étude possède plusieurs résultats intéressants, il est important de noter que les eQTL détectés dans le sang et dans les expectorations corrélaient peu. Parmi les 27 meilleurs *cis*-eQTL associés à la MPOC, seulement 5 étaient significatifs dans les deux tissus (18.5%). Cette observation confirme que le tissu employé pour réaliser des eQTL doit être pertinent pour le phénotype étudié.

CHAPITRE 2

Problématiques, objectifs et hypothèses des travaux

Cette thèse a pour objectif général d'approfondir nos connaissances sur la génétique de la MPOC, plus précisément 1) l'identification de nouveaux gènes et voies biologiques impliqués dans l'obstruction respiratoire, 2) l'identification de gènes causaux dans les régions chromosomiques précédemment associées à la MPOC, 3) l'identification de nouvelles régions chromosomiques impliquées dans le développement de la maladie et 4) l'amélioration du diagnostic pour le déficit en alpha-1 antitrypsine.

Voici une brève description de la problématique de chaque projet de recherche ainsi que les principaux objectifs et hypothèses de recherche.

Chapitre 3

Problématique

Dans les dernières années, plusieurs études pangénomiques ont identifié des régions chromosomiques impliquées dans le développement de la maladie[90,327–331]. Par contre, la majorité des SNP identifiés sont localisés dans des régions non codantes[397]. Une des explications possibles est que ces variations influencent le risque de développer la maladie en régulent l'expression de gènes.

Objectif

- Identifier des variations génétiques influençant à la fois l'expression de gènes et le risque de développer la MPOC et/ou de l'obstruction respiratoire.

Hypothèse

- Certaines variations génétiques vont influencer les chances de développer la MPOC en régulant l'expression de gènes dans les poumons.

Chapitre 4

Problématique

Le complexe majeur d'histocompatibilité est une région chromosomique associée avec des centaines de maladies[398]. Les gènes présents dans cette région sont des membres clés de la réponse immunitaire. Ce locus du génome humain est l'un des plus complexe à étudier avec son nombre élevé de gènes et son déséquilibre de liaison important[399]. Lors des dernières années, de nombreux GWAS ont associé des variations génétiques présentes dans cette région avec des maladies et phénotypes pulmonaires : asthme[400–405], cancer du poumon[375–377,406,407], fibrose kystique[408], pneumopathie interstitielle idiopathique[409], fonction pulmonaire[339,340,352,353]. Par contre, ces études n'ont pu déterminer les gènes causaux.

Objectifs

- Identifier des eQTL pulmonaires dans le complexe majeur d'histocompatibilité.
- Déterminer le ou les gènes régulés par les variations génétiques identifiées dans les GWAS précédents.

Hypothèses

- Un grand nombre d'eQTL pulmonaires sont présents dans le MHC.
- Des variations génétiques précédemment associées aux maladies respiratoires vont influencer les niveaux d'expression de gènes dans les poumons.
    - Certains des gènes régulés auront des fonctions biologiques pertinentes pour les phénotypes pulmonaires associés.

Chapitre 5

Problématique

La MPOC est une maladie complexe ayant comme principal facteur de risque le tabagisme. Par contre, seulement 15 à 25% de tous les fumeurs vont développer la maladie et ceux-ci n'auront pas tous les mêmes symptômes cliniques. De nombreuses évidences démontrent le rôle important de la génétique dans cette maladie[8]. Plusieurs études pangénomiques ont identifié des régions chromosomiques impliquées dans le développement de la maladie, mais les résultats obtenus n'expliquent qu'une faible proportion du risque. L'*International COPD Genetics Consortium* fut créé afin d'identifier de nouvelles régions de susceptibilité. Pour ce faire, de nombreuses cohortes de sujets MPOC ont été combinées afin de réaliser le plus grand GWAS sur la maladie.

Objectifs

- Identifier de nouvelles régions chromosomiques impliquées dans le développement de la MPOC.
- Déterminer les gènes possiblement causaux dans les loci identifiés.

Hypothèses

- Plusieurs régions du génome humain sont associées avec la MPOC.
    - Certaines de ces régions auront précédemment été associées avec la maladie et/ou d'autres phénotypes reliés.
    - De nouvelles régions seront identifiées grâce au nombre important d'échantillons employés.
- Des gènes avec des fonctions biologiques pertinentes seront identifiés en combinant les résultats du GWAS et d'eQTL pulmonaires.

Chapitre 6

Problématique

Le déficit en alpha-1 antitrypsine (DAAT) peut entraîner le développement d'emphysème chez les personnes non-fumeurs et causer l'emphysème précoce chez les fumeurs. En Amérique du Nord, il est estimé qu'une personne sur 3,000 à 5,500 souffre de cette condition[217]. Le déficit est causé par des mutations dans la séquence codante du gène *SERPINA1*. Le DAAT est sous-diagnostiqué par les cliniciens et la majorité des cas ne sont jamais identifiés. Les raisons expliquant cette situation sont nombreuses[220] : pénétrance incomplète, symptômes cliniques communs, diagnostique complexe. Le séquençage de l'ADN est déjà reconnu comme étant la meilleure technique, mais il est jugé dispendieux et difficile à implanter[45].

Objectifs

- Déterminer les variations génétiques présentes dans la séquence codante du gène *SERPINA1* dans une cohorte canadienne de patients MPOC.
- Évaluer la faisabilité et l'utilité du séquençage d'ADN du gène *SERPINA1* afin de diagnostiquer le déficit et d'implanter les recommandations des sociétés canadienne, américaine et européenne de thoracologie[45,220].

Hypothèses

- Une proportion de patients MPOC canadiens souffre de déficit en alpha-1 antitrypsine.
- Le séquençage de l'ADN permet de diagnostiquer le déficit rapidement avec une grande précision.

CHAPITRE 3

Genetic regulation of gene expression in the lung identifies CST3 and CD22 as potential causal genes for airflow obstruction

Lamontagne M*, Timens W*, Hao K*, Bossé Y, Laviolette M, Steiling K, Campbell JD, Couture C, Conti M, Sherwood K, Hogg JC, Brandsma CA, van den Berge M, Sandford A, Lam S, Lenburg ME, Spira A, Paré PD, Nickle D, Sin DD, Postma DS. Thorax. 2014 Nov;69(11):997-1004.

Genetic regulation of gene expression in the lung identifies CST3 and CD22 as potential causal genes for airflow obstruction

Maxime Lamontagne*[1], Wim Timens*[2], Ke Hao*[3], Yohan Bossé[1,4], Michel Laviolette[1], Katrina Steiling[5], Joshua D Campbell[5], Christian Couture[1], Massimo Conti[1], Karen Sherwood[6], James C Hogg[6,7], Corry-Anke Brandsma[2], Maarten van den Berge[8], Andrew Sandford[6,9], Stephen Lam[10], Marc E Lenburg[5], Avrum Spira[5], Peter D Paré[6,9], David Nickle[11], Don D Sin[6,9], Dirkje S Postma[†8]

1) Institut universitaire de cardiologie et de pneumologie de Québec, Québec, Canada;

2) University of Groningen, University Medical Center Groningen, Department of Pathology and Medical Biology, GRIAC research institute, Groningen, The Netherlands;

3) Department of Genetics and Genomics Sciences, Mount Sinai School of Medicine, New York, New York, USA;

4) Department of Molecular Medicine, Laval University, Québec, Canada;

5) Division of Computational Biomedicine, Bioinformatics Program, Boston University, Boston, Massachusetts, USA;

6) University of British Columbia Center for Heart Lung Innovation and Institute for Heart and Lung Health, St. Paul's Hospital, Vancouver, BC, Canada;

7) Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada;

8) University of Groningen, University Medical Center Groningen, Department of Pulmonology, GRIAC research institute, Groningen, The Netherlands;

9) Respiratory Division, Department of Medicine, The University of British Columbia, Vancouver, BC, Canada;

10) British Columbia Cancer Agency, Vancouver, British Columbia, Canada;

11) Merck & Co. Inc., Rahway, New Jersey, USA.

*These authors contributed equally

†Corresponding Author

Corresponding author:

Dirkje S. Postma, M.D., Ph.D.

University of Groningen,

University Medical Center Groningen,

Department of Pulmonary Medicine and Tuberculosis,

GRIAC Research Institute,

AA11, Hanzeplein, PO Box 30001, 9700 RB

Groningen, The Netherlands

Tel.: +31 (0)503613532; fax: +31 (0)503619320

Email: d.s.postma@umcg.nl

Keywords: Chronic Obstructive Pulmonary Disease, expression Quantitative Trait Loci, COPD Pathogenesis, Computational Biology, Lung Gene Expression

Word count: 3016

Running Head: Lung Gene Expression in COPD

Résumé

Introduction : La maladie pulmonaire obstructive chronique (MPOC) est une maladie complexe dont la pathogénèse est mal comprise. Les analyses intégratives ont le potentiel d'élucider les mécanismes impliqués dans le développement de la MPOC et des fonctions respiratoires. Récemment, nous avons combiné l'expression des gènes et les variations génétiques de 1,111 sujets pour identifier des eQTL pulmonaires.

Objectif : Déterminer des associations causales entre des variations génétiques associées à la MPOC et aux fonctions respiratoires et les changements dans l'expression des gènes dans les poumons.

Méthodes : Le lien entre les SNP et l'expression des gènes a été évalué pour trois phénotypes : % du VEMS prédit, le ratio VEMS/CVF et la MPOC. Différents modèles furent testés dans trois cohortes et les résultats ont été combinés dans une méta-analyse. Des analyses *in silico* ont évalué le rôle des voies biologiques pour les gènes identifiés. L'expression des gènes candidats a été validée dans un jeu de données de tissus pulmonaires et un de brossage bronchique.

Résultats : Des triplets SNP-ARNm-phénotype démontrant des relations causales ont été identifiés pour le pourcentage du VEMS prédit (n=169) et le ratio VEMS/CVF (n=80). Des variations génétiques associées à des fonctions respiratoires réduites régulaient à la hausse l'expression des gènes *CST3* et *CD22*, deux gènes avec des fonctions biologiques pertinentes. Les analyses des voies biologiques ont identifié une surreprésentation des gènes impliqués dans l'apoptose, le métabolisme des xénobiotiques, et les balances protéase-antiprotéase et oxydant-antioxydant.

Conclusion : En utilisant une approche intégrative pour analyser les liens entre les fonctions respiratoires, les SNP précédemment associés à la MPOC et l'expression des gènes dans les poumons, nous avons identifié les gènes *CST3* et *CD22* comme gènes potentiellement causaux pour le développement d'obstruction respiratoire.

Abstract

Background: Chronic obstructive pulmonary disease (COPD) is a complex chronic disease with poorly understood pathogenesis. Integrative genomic approaches have the potential to elucidate the biological networks underlying COPD and lung function. We recently combined genome-wide genotyping and gene expression in 1,111 human lung specimens to map expression Quantitative Trait Loci (eQTL).

Objective: To determine causal associations between COPD and lung function-associated single nucleotide polymorphisms (SNPs) and lung tissue gene expression changes in our lung eQTL dataset.

Methods: We evaluated causality between SNPs and gene expression for three COPD phenotypes: FEV1%predicted, FEV1/FVC, and COPD as a categorical variable. Different models were assessed in the three cohorts both independently and in a meta-analysis. SNPs associated with a COPD phenotype and gene expression were subjected to causal pathway modeling and manual curation. In silico analyses evaluated functional enrichment of biological pathways amongst newly identified causal genes. Biologically relevant causal genes were validated in two separate gene expression datasets of lung tissues and bronchial airway brushings.

Results: High reliability causal relations were found in SNP-mRNA-phenotype triplets for FEV1%predicted (n=169) and FEV1/FVC (n=80). Several genes of potential biological relevance for COPD were revealed. eQTL-SNPs up-regulating CST3 and CD22 were associated with worse lung function. Signaling pathways enriched with causal genes included xenobiotic metabolism, apoptosis, protease-antiprotease and oxidant-antioxidant balance.

Conclusion: By using integrative genomics and analyzing the relationships of COPD phenotypes with SNPs and gene expression in lung tissue, we have identified CST3 and CD22 as potential causal genes for airflow obstruction. This study also augmented the understanding of previously described COPD pathways.

Key messages

What is the key question?

What are the causal genetic variants changing gene expression in the lung that in turn associate with lower lung function and COPD?

What is the bottom line?

Lung function-associated genetic variants alter the mRNA expression of nearby genes involved in biological pathways underpinning pulmonary function and COPD pathogenesis.

Why read on?

New genes of airflow obstruction are identified with a generalized framework for the identification of causal genes from joint examination of genome-wide genotyping and gene expression data in the same patients.

Introduction

Genome-wide association studies (GWAS) have revolutionized our ability to identify common genetic variants that are associated with complex chronic diseases[1]. This approach has been applied to chronic obstructive pulmonary disease (COPD), a lung disease that is caused predominantly by cigarette smoking in the western world[2-4]. It is well-known that only a subset of heavy smokers (15-20%) develops clinically relevant COPD and there is considerable evidence that there is a substantial genetic component involved in its pathogenesis[5]. Although GWAS have identified novel loci that harbor susceptibility genes, they do not allow precise identification of the causal variant (or variants). In addition, GWAS do not provide information on how and to what extent the gene (or genes) within the susceptibility loci contribute to the phenotype. Interestingly, the majority of genetic variants which have been associated with disease traits by GWAS do not affect the coding sequence of genes but are located in intergenic regions or introns[6]. Possible explanations are that the associated alleles are in linkage disequilibrium (LD) with rarer coding alleles with large effect sizes[7] and/or that the genetic variants control the level of expression of genes involved in pathogenetic pathways. For complex genetic diseases, such as COPD, the effects of susceptibility alleles may primarily act by regulating gene expression rather than by altering protein coding as in most Mendelian diseases[8].

We recently reported the discovery of a large number of lung-specific expression quantitative trait loci (eQTLs)[9] and identified the most likely causal genes within three GWAS-nominated COPD susceptibility loci[10]. The aim of the present study is to use the power of genome-wide mRNA expression arrays combined with genome-wide interrogation of single nucleotide polymorphisms (SNPs) to pinpoint specific SNPs that are related to lung tissue gene expression and to COPD phenotypes. Identification of susceptibility alleles that function as strong eQTLs increases the likelihood of identifying the true susceptibility gene within loci with large areas of LD[8]. Moreover, the use of integrative genomics by combining environmental exposure data with susceptibility alleles, RNA expression levels, and different COPD phenotypes can unravel causal genetic relationships[11]. The basic principle of the current study is to map the genetic regulation of gene expression to identify DNA variants that induce changes in transcriptional networks that in turn contribute to COPD and airway obstruction pathogenesis.

Methods

Subject selection

The methods for subject selection and phenotyping as well as for interrogation of gene expression and genotype were recently described[12]. The lung tissue used for discovery of eQTLs was from 1,111 human subjects who underwent lung surgery at three academic sites, Laval University, University of British Columbia (UBC), and University of Groningen, henceforth referred to as Laval, UBC, and Groningen, respectively. All lung specimens from Laval were obtained from patients undergoing lung cancer surgery and were harvested from a site distant from the tumor. At UBC, the majority of samples were from patients undergoing resection of small peripheral lung lesions. Additional samples were from autopsy and at the time of lung transplantation. At Groningen, the lung specimens were obtained at surgery from patients with various lung diseases, including patients undergoing therapeutic resection for lung tumors, harvested from a site distant from the tumor, and lung transplantation. For the present study, the principal aim was to examine smoking-related airway obstruction. Thus, we excluded subjects whose lung function may have been influenced by lung diseases other than COPD and lung cancer. Exclusion criteria are provided in online supplement.

COPD phenotypes and genome-wide association study

Genome-wide association was performed using linear or logistic regression models on the three phenotypes: FEV_1% predicted and FEV_1/FVC as continuous variables, and COPD defined dichotomously based on an $FEV_1/FVC < 0.7$ cut-off (see online supplement). Single-marker association tests were run within each cohort adjusting for age, gender and smoking status. Fixed-effects meta-analysis was then performed combining the three cohorts using inverse standard error weighting.

Expression trait processing

Expression traits were adjusted for age, gender, and smoking status as described previously[9]. Gene expression data are available in the Gene Expression Omnibus repository through accession number GSE23546.

Causality Models

We evaluated three competing causality models to describe the relationship between lung eQTL-SNPs, RNA expression and COPD phenotypes[12] (**Figure 3-1**). Model 1 indicates a potential causal relationship where the SNP acts on gene expression to produce the phenotype, which was the main interest of this work. Model 2 indicates that the gene expression pattern is reactive to the phenotype, i.e. the phenotype drives gene expression. Model 3 is the independent model where the SNP acts on the phenotype and gene expression independently. We required a p-value $< 1 \times 10^{-3}$ for SNP associated with phenotype before examining a particular variable triplet (i.e. SNP, gene expression, and phenotypes). The causal model (Model 1) was selected when p-value < 0.05 was found for SNP associated with gene expression adjusting for phenotype and p-value > 0.05 was found for SNP associated with phenotype adjusting for gene expression. More details are provided in the online supplement. Analyses were performed in the three cohorts separately and then combined into a meta-analysis. We then conducted sample bootstrapping and repeated the causality test for N=1000 realizations. The reliability score of the molecular relationship is the fraction of bootstrap realizations that supports the call observed. Threshold for p-values, number of bootstrap reselections, and reliability cut-off (>0.8) were selected based on previous literature[12].

Manual curation and pathway analyses on reliable causal models

The biology and possible role of genes in the causal model were reviewed by manual curation and bioinformatics tools including Ingenuity Pathways Analysis (IPA), MetaCore, and Partek (see online date supplement).

Replication studies

Biologically relevant causal genes were validated in two datasets. First, a bronchial airway epithelial dataset where genome-wide gene expression levels were obtained from bronchial brushing of 238 individuals and associated with lung function as previously described[13]. Second, a regional lung tissue dataset where genome-wide gene expression were obtained for eight regions of the same lung and for eight patients. The association

between gene expression levels and micro-CT based regional emphysema severity was assessed[14]. More details are provided in the online data supplement.

Results

Of the 1,111 subjects in whom there were data on genotype and gene expression, 848 with sufficient phenotypic information were included in the analysis. The demographic and clinical features of the subjects in the three cohorts are described in **Table 3-1**.

Causal pathways that fit model 1

Model fitting was restricted to SNPs that were significantly associated with at least one of the lung function variables or COPD as a categorical variable at a p-value cut-off of 10^{-3}. SNPs were considered an eQTL if they had a p-value $\leq 10^{-5}$. Using these criteria, there were 4,465 SNP-mRNA-phenotype triplets. Of these, 249 triplets showed a significant fit to the causal model for at least 1 of the phenotypes with a reliability score > 0.8 in at least one cohort and/or in the meta-analysis. A fit to this model means that the SNP was associated with one of the phenotypes and affected gene expression in a direction that supported a causal relationship.

The list of causal models with a confidence score > 0.8 in any one of the cohorts or in the meta-analysis is shown in **Supplementary Table 3-S1**. For FEV_1% predicted, 169 causal models were found including 122 unique SNPs and 169 probe sets. The 169 models included 168 *cis* and 1 *trans*-eQTLs, respectively. For FEV_1/FVC, 80 causal models were found involving 63 unique SNPs and 80 probe sets. Among these 80 models, 79 were *cis*-eQTLs and 1 was *trans*-eQTLs. No causal model was found when using COPD as a categorical variable. There was no overlap between causal pathway models for FEV_1% predicted and FEV_1/FVC.

Manual curation of significant and reliable causal models

Significant causal models were inspected manually. We identified a number of models of potential biological relevance for FEV_1% predicted and FEV_1/FVC. The most

biologically relevant models are listed in **Table 3-2**. The p-values and direction of effect for each of the SNP association with the gene expression and the phenotype are also indicated. For example, SNP rs6048956 was significantly associated with cystatin C (*CST3*) transcript. The common allele was associated with higher expression of the transcript (**Figure 3-2** and positive eQTL Z score in **Table 3-2**) and with lower FEV_1 % predicted **(Figure 3-2** and negative Z score with phenotype in **Table 3-2**). Taken together, these results suggest that the common allele confers susceptibility to a lower FEV_1 % predicted value through up-regulation of the CST3 mRNA expression levels in the lung. This was confirmed in a second causality model that interrogated a different probe set for CST3, FEV_1 % predicted and rs6515375 (**Table 3-2** & **Supplementary Figure 3-S1**). Analogous observations were made for natural cytotoxicity triggering receptor 3 (*NCR3*), cystatin A (*CSTA*), peroxisome proliferation-activated receptor gamma coactivator 1 alpha (*PPARGC1A*), folliculin (*FLCN*), BCL2-like 1 (*BCL2L1*), cell adhesion molecule 2 (*CADM2*), and tumor necrosis factor receptor superfamily member 10b (*TNFRSF10B*) for FEV_1 % predicted, and glutathione S-transferase omega 2 (*GSTO2*), DEP domain containing 6 (*DEPDC6*), CD22 molecule (*CD22*), mucin 22 (*MUC22*), and serine peptidase inhibitor Kazal type 5 (*SPINK5*) for FEV_1/FVC (**Table 3-2**). The direction of effects for all these causality models is illustrated in **Supplementary Figures S2-S13**.

Pathway analyses

To find key biological pathways involved in COPD, causality genes were overlaid on canonical pathways available in IPA. 169 and 80 causality genes for FEV_1 and FEV_1/FVC were considered, respectively (**Supplementary Table 3-S1**). **Table 3-3** shows all canonical pathways enriched with causality genes ($p < 0.05$). Of interest was the Aryl Hydrocarbon Receptor Signaling pathway, which is involved in xenobiotic clearance. Five causality genes were noted in this canonical pathway including *ARNT* (aryl hydrocarbon receptor nuclear translocator), *IL6* (interleukin 6), *GSTO2* (glutathione S-transferase omega 2), *ALDH8A1* (aldehyde dehydrogenase 8 family, member A1), and *TRIP11* (thyroid hormone receptor interactor 11). The Aryl Hydrocarbon Receptor Signaling pathway and the location of the causality genes are illustrated in **Supplementary Figure 3-S14**. Another pathway involved in xenobiotic handling, the Xenobiotic Metabolism Signaling pathway,

was also enriched for causality genes, many of which overlapped with those in the Aryl Hydrocarbon Receptor Signaling pathway (*ARNT*, *IL6*, *GSTO2*, and *ALDH8A1*). The former also includes *PPARGC1A* (peroxisome proliferator-activated receptor gamma, coactivator 1 alpha), and *PRKCE* (protein kinase C, epsilon). This pathway is illustrated in **Supplementary Figure 3-S15**.

For the Partek analyses 118 and 45 transcripts were identified for FEV_1 % predicted and FEV_1/FVC, respectively. These transcripts mapped to a total of 115 genes in Partek GS (based on official gene symbols) and were parsed to Partek Pathway Suite. Causality genes were overlaid onto canonical pathways from the REACTOME and KEGG databases. **Table 3-4** shows the canonical pathways enriched with causality genes ($p < 0.05$). TNF-related apoptosis-inducing ligand (TRAIL) signaling was the most significant pathway with several other apoptosis related pathways include in the list of top significant associations. The glutathione metabolism pathway connects to both the cyanoamino acid and taurine and hypotaurine metabolism pathways (**Supplementary Figure 3-S16**). These three pathways were identified in our enrichment analysis and were also identified using the IPA system. GO functional categories also showed enrichment for causality genes **(Supplementary Figure 3-S17**) including the gamma-glutamyl transferase (GGT) activity.

Replication of the most biologically relevant causality models

No other large-scale lung eQTL dataset is available in patients with and without COPD. To replicate the most biologically relevant causality model, we relied on two genome-wide expression datasets. Causality genes were first validated in a gene expression study of bronchial airway epithelial cells obtained by bronchoscopy. All fourteen genes represented in **Table 3-2**, except MUC22, were assayed in this airway dataset. Two genes were significantly associated with lung function at an FDR of 5% in the bronchial airway epithelial dataset: CSTA (FDR = 0.002 with FEV_1 % predicted) and TNFRSF10B (FDR = 9.35×10^{-5} with FEV_1 % predicted). In both cases, the direction of effect was the same as that in the current study (**Table 3-5**). For example, higher CSTA mRNA levels were associated with worse lung function (**Table 3-2** and **Supplementary Figure 3-S3**).

Causality genes identified in this study were also compared to a second lung transcriptomic study that evaluated the impact of regional emphysema severity on gene expression. Interestingly, CD22 was positively associated with regional emphysema severity within individuals ($p = 5.8 \times 10^{-5}$) and was a part of the 127 gene signature for emphysema identified in that study[14]. This observation is consistent with the current study showing that carriers of the rare allele for rs10411704 had greater mRNA expression of CD22 and worse FEV_1/FVC (**Table 3-2** and **Supplementary Figure 3-S11**). Associations with regional emphysema severity were also observed for three other genes in **Table 3-2** including *NCR3* (p=0.058), *PPARGC1A* (p=0.081), and *BCL2L1* (p=0.051), but these were not statistically significant. However, the direction of effect was consistent only for PPARGC1A. The expression of this gene was found to decrease with emphysema severity in the regional lung tissue dataset and, in the current study, carriers of the rare allele for rs4550905 were associated with less mRNA expression of PPARGC1A in the lung and lower FEV_1 (**Table 3-2** and **Supplementary Figure 3-S4**).

Discussion

This investigation integrated a genome-wide eQTL study on lung tissue with genome-wide genetic association results for lung function and COPD in the same subjects. For the discovery of eQTLs we used the entire dataset, which consisted of 1,111 tissue samples from subjects who had lung surgery for a variety of reasons. In order to focus on COPD, we limited the study to 848 subjects with sufficient phenotypic information and without a lung disease (other than COPD and lung cancer), which could cause abnormalities of pulmonary function. We limited the causal pathway analysis to SNPs that were both significantly related to gene expression ($p < 10^{-5}$) and were associated with one of three COPD phenotypes ($p < 10^{-3}$), i.e. FEV_1 % predicted, FEV_1/FVC and COPD defined as $FEV_1/FVC < 0.7$. Of the 4,465 SNP-mRNA-phenotype triplets which met our inclusion criteria, 249 triplets showed a significant fit to the causality model for at least 1 of the phenotypes with a reliability score ≥ 0.8. We thus provide evidence that these SNPs influence disease susceptibility by altering gene expression. Causality pathway genes were

enriched in pathways involved in xenobiotic handling, antiprotease and antioxidant activity and apoptosis.

Two of the eQTL-SNPs in *CST3* (rs6515375 and rs6048956) and another for *CSTA* (rs2270859) were in the causal pathway for FEV_1 % predicted. The two *CST3* SNPs were in perfect LD and were eQTLs for two different probe sets. *CST3* and *CSTA* are cysteine antiproteases; *CST3* antagonizes cysteine cathepsins such as Cathepsin L and S, and *CSTA* acts similarly for cathepsins B, H and L[15,16]. Interestingly the direction of association is such that the alleles that are associated with a higher mRNA level of cystatin C and cystatin A are associated with lower FEV_1/FVC.

The above findings might seem paradoxical since cystatins are antiproteases and if one simply invoked the protease-antiprotease hypothesis, then one might expect that individuals with higher levels to be protected from COPD. One possibility is that upregulation of *CST3* mRNA and protein could be the result of a feedback loop stimulated by high protease levels. This would be supported by observations in bronchoalveolar lavage fluid and serum reporting higher cystatin C level in patients with emphysema[17, 18]. Alternatively excessive inhibition of proteases may confer biological effects on a yet to be discovered mechanism which contributes to the pathogenesis of airflow obstruction. Further molecular studies of the involved proteins might shed more light on this. We previously reported SNPs associated with *CST3* mRNA and protein levels in alveolar macrophages[19]. However, in the latter study, higher *CST3* mRNA in alveolar macrophages was associated with higher FEV_1. Together, these studies suggest that the cystatin genes are important in the pathogenesis of COPD, however, the relationship between *CST3* mRNA expression levels and lung function remains to be elucidated in relevant tissues and cell types.

With respect to anti-oxidant activity, eQTLs from *PPARGC1A* and *GSTO2* were found. Peroxisome proliferator-activated receptor-gamma coactivator-1alpha (*PPARGC1A*) binds to PPARγ by induction of PPARγ ligands, co-activating PPARγ target genes, involved in anti-oxidant activity. Compared with controls, expression levels of PPARγ, PGC-1α, and γ-GCS have been reported to be significantly increased in the lungs of patients with mild COPD, and progressively decreased in more severe disease[20]. These

authors concluded that γ-GCS showed compensatory upregulation in the early stage of COPD, which progressively decompensated with disease progression and that the activation of the PPARγ/PGC-1α pathway may protect against COPD progression by upregulating γ-GCS and relieving oxidative stress. Furthermore *PPARGC1A* has been described to be involved in bronchial smooth muscle remodelling and skeletal muscle wasting[21, 22]. For Glutathione-S-transferase Omega (*GSTO*) 2, a strong association of the Asn142Asp SNP with FEV_1 and FVC was found in the Framingham Heart Study[23]. In a latter study, the Asn142Asp polymorphism in *GSTO2* and the *GSTO1* 140Asp/*GSTO2* 142Asp haplotype were associated with increased risk of COPD but failed to reveal an association between lung function parameters and non-synonymous coding SNPs in the *GSTO* genes[24]. Polymorphisms in *GSTO2* were also associated with COPD either with or without lung cancer[25].

Pathways involved in apoptosis and in handling of xenobiotic pathways were significantly enriched for causal genes. The results of the present study aid in understanding the underlying genetic (dys)regulation of apoptosis. In general, increased apoptosis of endothelial cells and fibroblasts has been shown to contribute to the development of emphysema[26, 27]. This is partly due to an imbalance caused by excess oxidant and protease effects related to cigarette smoking and to intrinsic dysregulation of apoptosis induced in susceptible individuals. This may explain why the apoptotic effects are larger in smokers with COPD than smokers without COPD. Dysregulation of apoptosis can also work the other way: decreased apoptosis in cells of the immune system can lead to sustained inflammation, and when occurring in fibroblasts (e.g. in the bronchial wall) can induce fibrosis. Emphysema is characterized by alveolar cell apoptosis, which was shown to be mediated by increased levels of apoptotic proteins including TRAIL receptors[28]. Interestingly, TRAIL signaling was the most significant pathway enriched with causality genes using Partek. TRAIL signaling has been shown to regulate immune responses in the lung and can lead to a sustained, pro-inflammatory response that contributes to vascular disease. An opposite effect is related to the Death Receptor signaling pathway, which, in human, binds with TRAIL, induces formation of a death-inducing signaling complex ultimately leading to caspase activation and initiation of apoptosis[29].

The Aryl Hydrocarbon Receptor Signaling pathway is involved in xenobiotic clearance and therefore of relevance to the known vulnerability of COPD patients to (cigarette) smoke. Furthermore additional metabolic pathways including the glutathione metabolism pathway and two subpathways involving metabolism of the cyanoamino acids, taurine and hypotaurine, were also significantly enriched in the causal analysis. Genetic variants in a number of genes in the glutathione pathway have previously been associated with risk for COPD[25,30].

No other large-scale lung eQTL dataset is available in patients with and without COPD. To replicate the most biologically relevant causality models, we relied on two genome-wide expression datasets. Causality genes were first validated in a gene expression study of bronchial airway epithelial cells obtained by bronchoscopy and then in a transcriptomic study of whole lung explanted at surgery. Whole genome genotyping is not available for these datasets. Accordingly, replication of significant triplets (i.e. SNP-mRNA-phenotype) can only evaluate the concordance between gene expression and phenotype. Additional studies with phenotype, genotype and gene expression in the lung would be required to provide full validation.

In conclusion integration of lung-specific eQTL data with GWAS from the same individuals has revealed interesting and potentially causal pathways of airflow obstruction. GWAS have revolutionized our ability to identify gene variants that contribute to susceptibility for common complex genetic diseases but often do not pinpoint the exact genes or mechanisms. Causal pathway analysis involving the joint examination of genetic and genomic data is a vital next step in discovering novel biomarker and therapeutic targets in airflow obstruction as evidenced in the present study.

Acknowledgments

The authors would like to thank Christine Racine and Sabrina Biardel at the IUCPQ site of the Respiratory Health Network Biobank of the FRQS for their valuable assistance. We also acknowledge the staff at Calcul Québec for IT support with the high performance computer clusters. At UBC we thank the biobank coordinator Dr Mark Elliott. At the Groningen UMCG site Marnix Jonker is thanked for his support in selecting, handling and sending of lung tissues. We thank Dr. Joshua Millstein of University of Southern California for helpful advice on causality inference methodology.

Contributorship

WT, KH, YB, MLaviolette, PDP, DDS and DSP were involved in the conception and design of the study. MLamontagne, WT, KH, YB, MLaviolette, KSteiling, JDC, CC, MC, KSherwood, JCH, CAB, SL, MLenburg, AS, PDP, DN and DSP were involved in the acquisition and/or the analysis and interpretation of the data. MLamontagne, WT, KH, YB, MB, AS, PDP, DDS and DSP contributed in the writing or the revision of the manuscript.

Competing interests

CAB received a grant from Rosetta Merck. DN is a full time employee of Merck. DSP received consultancy fees from AstraZeneca, Boehringer Ingelheim, Chiesi, GlaxoSmithKline, Takeda, TEVA and a grant from Chiesi. DDS has served on advisory boards of Almirall, Nycomed, Talecris, AstraZeneca, Merck Frosst, Novartis and GlaxoSmithKline, received grants from AstraZeneca, GlaxoSmithKline and Wyeth, and received honoraria for speaking engagements from Takeda, AstraZeneca, GlaxoSmithKline and Boehringer Ingelheim. JDC received consultancy fees from Metera Biosciences and immuneering. MB received grants from TEVA, AstraZaneca and Chiesi. WT reveived a grant from Merck Sharp Dohme, received consultancy fees from Pfizer, and received lecture fees from GlaxoSmithKline, Chiesi and Roche Diagnosis.

Funding

This study was funded by Merck Research Laboratories, the Chaire de pneumologie de la Fondation JD Bégin de l'Université Laval, the Fondation de l'Institut universitaire de cardiologie et de pneumologie de Québec, the Respiratory Health Network of the FRQS, the Cancer Research Society and Read for the Cure, and the Canadian Institutes of Health Research (MOP – 123369). Y. Bossé was a research scholar from the Heart and Stroke Foundation of Canada and he is now recipient of a Junior 2 Research Scholar award from the Fonds de recherche Québec – Santé (FRQS). Don Sin is a Tier 1 Canada Research Chair for COPD.

References

1. Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med* 2010;363:166-76.

2. Cho MH, Boutaoui N, Klanderman BJ et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet* 2010;42:200-2.

3. Cho MH, Castaldi PJ, Wan ES et al. A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Hum Mol Genet* 2012;21:947-57.

4. Pillai SG, Ge D, Zhu G et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* 2009;5:e1000421.

5. Bosse Y. Updates on the COPD gene list. *Int J Chron Obstruct Pulmon Dis* 2012;7:607-31.

6. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science* 2008;322:881-8.

7. Thun GA, Imboden M, Ferrarotti I et al. Causal and Synthetic Associations of Variants in the SERPINA Gene Cluster with Alpha1-antitrypsin Serum Levels. *PLoS Genet* 2013;9:e1003585.

8. Nicolae DL, Gamazon E, Zhang W et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet* 2010;6:e1000888.

9. Hao K, Bosse Y, Nickle DC et al. Lung eQTLs to Help Reveal the Molecular Underpinnings of Asthma. *PLoS Genet* 2012;8:e1003029.

10. Lamontagne M, Couture C, Postma DS et al. Refining Susceptibility Loci of Chronic Obstructive Pulmonary Disease with Lung eqtls. *PLoS One* 2013;8:e70220.

11. Sieberts SK, Schadt EE. Moving toward a system genetics view of disease. Mammalian genome : official journal of the International Mammalian *Genome Society* 2007;18:389-401.

12. Schadt EE, Lamb J, Yang X et al. An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet* 2005;37:710-7.

13. Steiling K, van den Berge M, Hijazi K et al. A dynamic bronchial airway gene expression signature of chronic obstructive pulmonary disease and lung function impairment. *Am J Respir Crit Care Med* 2013;187:933-42.

14. Campbell JD, McDonough JE, Zeskind JE et al. A gene expression signature of emphysema-related lung destruction and its reversal by the tripeptide GHK. *Genome Med* 2012;4:67.

15. Butler MW, Fukui T, Salit J et al. Modulation of cystatin A expression in human airway epithelium related to genotype, smoking, COPD, and lung cancer. *Cancer Res* 2011;71:2572-81.

16. Pavlova A, Bjork I. Grafting of features of cystatins C or B into the N-terminal region or second binding loop of cystatin A (stefin A) substantially enhances inhibition of cysteine proteinases. *Biochemistry* 2003;42:11326-33.

17. Rokadia HK, Agarwal S. Serum cystatin C and emphysema: results from the National Health and Nutrition Examination Survey (NHANES). *Lung* 2012;190:283-90.

18. Takeyabu K, Betsuyaku T, Nishimura M et al. Cysteine proteinases and cystatin C in bronchoalveolar lavage fluid from subjects with subclinical emphysema. *Eur Respir J* 1998;12:1033-9.

19. Ishii T, Abboud RT, Wallace AM et al. Alveolar macrophage proteinase/antiproteinase expression and lung function/emphysema. *Eur Respir J* 2013.

20. Li J, Dai A, Hu R et al. Positive correlation between PPARgamma/PGC-1alpha and gamma-GCS in lungs of rats and patients with chronic obstructive pulmonary disease. *Acta Biochim Biophys Sin (Shanghai)* 2010;42:603-14.

21. Trian T, Benard G, Begueret H et al. Bronchial smooth muscle remodeling involves calcium-dependent enhanced mitochondrial biogenesis in asthma. *J Exp Med* 2007;204:3173-81.

22. Remels AH, Gosker HR, Schrauwen P et al. TNF-alpha impairs regulation of muscle oxidative phenotype: implications for cachexia? *FASEB J* 2010;24:5052-62.

23. Wilk JB, Walter RE, Laramie JM et al. Framingham Heart Study genome-wide association: results for pulmonary function measures. *BMC Med Genet* 2007;8 Suppl 1:S8.

24. Yanbaeva DG, Wouters EF, Dentener MA et al. Association of glutathione-S-transferase omega haplotypes with susceptibility to chronic obstructive pulmonary disease. *Free Radic Res* 2009;43:738-43.

25. de Andrade M, Li Y, Marks RS et al. Genetic variants associated with the risk of chronic obstructive pulmonary disease with and without lung cancer. *Cancer Prev Res (Phila)* 2012;5:365-73.

26. Park JW, Ryter SW, Kyung SY et al. The phosphodiesterase 4 inhibitor rolipram protects against cigarette smoke extract-induced apoptosis in human lung fibroblasts. *Eur J Pharmacol* 2013;706:76-83.

27. Yang Q, Underwood MJ, Hsin MK et al. Dysfunction of pulmonary vascular endothelium in chronic obstructive pulmonary disease: basic considerations for future drug development. *Curr Drug Metab* 2008;9:661-7.

28. Morissette MC, Vachon-Beaudoin G, Parent J et al. Increased p53 level, Bax/Bcl-x(L) ratio, and TRAIL receptor expression in human emphysema. *Am J Respir Crit Care Med* 2008;178:240-7.

29. Bodmer JL, Holler N, Reynard S et al. TRAIL receptor-2 signals apoptosis through FADD and caspase-8. *Nat Cell Biol* 2000;2:241-3.

30. He JQ, Connett JE, Anthonisen NR et al. Glutathione S-transferase variants and their interaction with smoking on lung function. *Am J Respir Crit Care Med* 2004;170:388-94.

Tables

Table 3-1 : Clinical characteristics of patients.

Characteristics	Laval (n=403)	UBC (n=270)	Groningen (n=175)
Age (years)	63.4 ± 9.8	63.9 ± 10.0	59.7 ± 10.0
Male/female (n)	224/179	144/126	83/92
Body Mass Index (kg/m^2)	26.6 ± 5.2	25.7 ± 5.4	24.6 ± 4.11
FEV_1% Predicted	82.19 ± 17.6	83.6 ± 22.3	72.4 ± 24.9
FEV_1/FVC	0.68 ± 0.10	0.69 ± 0.13	0.64 ± 0.16
COPD (%)	209 (51.9)	114 (42.2)	112 (64.0)
Stage 1 : Mild	80 (38.2)	43 (37.7)	16 (14.3)
Stage 2 : Moderate	117 (56.0)	60 (52.6)	30 (26.8)
Stage 3 : Severe	11 (5.3)	2 (1.8)	11 (9.8)
Stage 4 : Very Severe	1 (0.5)	9 (7.9)	44 (39.3)
Missing data	0	0	11 (9.8)
Non-COPD (%)	194 (48.1)	156 (57.8)	63 (36.0)
Smoking (%)			
Smoker	89 (22.1)	91 (33.7)	12 (6.9)
Ex-Smoker	281 (69.7)	150 (55.6)	117 (66.9)
Non-Smoker	33 (8.2)	16 (5.9)	43 (24.6)
Missing data	0	13 (4.8)	3 (1.7)
Pack years (years)	48.6 ± 27.4	46.0 ± 28.6	32.1 ± 18.1

Continuous variables are presented as means ± standard deviations.

Table 3-2 : Models of biological relevance identified by manual curation.

SNPs	Reference allele (freq)[a]	Gene (probe set)	Lung eQTL p-value [b]	eQTL Z score[c]	p-value phenotype [d]	Z score phenotype [e]
FEV1 % predicted						
rs769178	G (0.91)	*NCR3* (100125842_TGI_at)	8.03×10^{-30}	11.3	4.7×10^{-4}	3.5
rs6048956	C (0.77-0.79)	*CST3* (100307577_TGI_at)	1.54×10^{-68}	17.5	7.8×10^{-4}	-3.4
rs6515375	G (0.79-0.81)	*CST3* (100125967_TGI_at)	3.65×10^{-6}	4.6	5.8×10^{-4}	-3.4
rs2270859	G (0.83-0.88)	*CSTA* (100148334_TGI_at)	1.74×10^{-9}	6.0	3.1×10^{-6}	-4.7
rs4550905	G (0.26-0.31)	*PPARGC1A* (100131093_TGI_at)	1.69×10^{-5}	-4.3	4.4×10^{-4}	-3.5
rs3803761	G (0.71-0.77)	*FLCN* (100135396_TGI_at)	3.77×10^{-29}	11.2	3.4×10^{-2}	-2.1
rs1543438	A (0.76-0.82)	*BCL2L1* (100158784_TGI_at)	3.89×10^{-5}	-4.1	7.94×10^{-5}	3.9
rs9880397	G (0.62-0.64)	*CADM2* (100162763_TGI_at)	1.3×10^{-8}	-5.7	2.9×10^{-4}	3.6
rs2466183	T (0.84-0.86)	TNFRSF10B (100153254_TGI_at)	3.76×10^{-5}	-4.1	6.6×10^{-4}	3.4
FEV_1/FVC						
rs17754977	A (0.30-0.32)	*GSTO2* (100132911_TGI_at)	3.96×10^{-5}	4.1	8.2×10^{-4}	3.3
rs9987135	T (0.28-0.31)	*DEPDC6* (100154484_TGI_at)	6.13×10^{-59}	16.2	1.1×10^{-4}	-3.9
rs10411704	T (0.79-0.82)	*CD22* (100154732_TGI_at)	1.70×10^{-40}	-13.3	4.6×10^{-4}	3.5
rs12179536	A (0.80-0.84)	*MUC22* (100304000_TGI_at)	3.47×10^{-33}	-12.0	3.0×10^{-3}	3.0
rs2287765	T (0.91-0.93)	*SPINK5* (100305138_TGI_at)	2.85×10^{-10}	6.3	9.2×10^{-4}	-3.3

P-values and Z scores in this table are from the meta-analysis. As indicated in the text, the significant causal models were selected based on results of both individual cohorts and meta-analysis (SNP associated with phenotype with p-value $< 10^{-3}$, SNPs associated with gene expression with p-value $\leq 10^{-5}$, and reliability score > 0.8 in at least one cohort and/or in the meta-analysis).

[a]Frequency of the reference allele in the three cohorts.

[b]Lung eQTL p-value from the meta-analysis.

[c]Z score from the eQTL meta-analysis showing the direction of effect for the SNP on gene expression.

[d]P-value for association between the SNP and phenotype from the meta-analysis.

[e]Effect size (Z score) for association between the SNP and the phenotype from the meta-analysis, showing the direction of effect for the SNP on the phenotype.

Known role of genes are provided in **Supplementary Table S2**.

Table 3-3 : Canonical pathways enriched for causality genes using the Ingenuity Pathway Analysis system.

Pathways	P-value
Cdc42 Signaling	0.00002
Crosstalk between Dendritic Cells and Natural Killer Cells	0.00015
Graft-versus-Host Disease Signaling	0.00069
OX40 Signaling Pathway	0.00151
Antigen Presentation Pathway	0.00468
Neuroprotective Role of THOP1 in Alzheimer's Disease	0.00589
Taurine and Hypotaurine Metabolism	0.00617
Autoimmune Thyroid Disease Signaling	0.00661
Regulation of Actin-based Motility by Rho	0.00676
Communication between Innate and Adaptive Immune Cells	0.00676
Aryl Hydrocarbon Receptor Signaling	0.00759
Integrin Signaling	0.00891
Allograft Rejection Signaling	0.00977
Cytotoxic T Lymphocyte-mediated Apoptosis of Target Cells	0.01023
Cyanoamino Acid Metabolism	0.01047
Glutathione Metabolism	0.01072
Amyotrophic Lateral Sclerosis Signaling	0.01230
Systemic Lupus Erythematosus Signaling	0.01230
PXR/RXR Activation	0.01995
Macropinocytosis Signaling	0.02455
Selenoamino Acid Metabolism	0.02692
Xenobiotic Metabolism Signaling	0.02951
Cellular Effects of Sildenafil (Viagra)	0.03162
Aminoacyl-tRNA Biosynthesis	0.03631
Actin Cytoskeleton Signaling	0.04365
FAK Signaling	0.04467
Apoptosis Signaling	0.04677

Pathways discussed in the text are given in italic.

Table 3-4 : Canonical pathways enriched for causality genes using Partek.

Pathways	P-value	Pathway ID
TRAIL signaling	0.0003	reactome_pathway_506
Cyanoamino acid metabolism	0.001	kegg_pathway_36
Cell adhesion molecules (CAMs)	0.002	kegg_pathway_139
Extrinsic Pathway for Apoptosis	0.002	reactome_pathway_502
Death Receptor Signaling	0.002	reactome_pathway_503
Taurine and hypotaurine metabolism	0.002	kegg_pathway_34
Glutathione metabolism	0.003	kegg_pathway_39
Aminoacyl-tRNA biosynthesis	0.004	kegg_pathway_81
Apoptosis	0.006	reactome_pathway_501
Cytosolic tRNA aminoacylation	0.006	reactome_pathway_1127
Glutathione conjugation	0.007	reactome_pathway_256
Natural killer cell mediated cytotoxicity	0.011	kegg_pathway_152
The NLRP1 inflammasome	0.015	reactome_pathway_469
tRNA Aminoacylation	0.018	reactome_pathway_1126
Apoptosis	0.023	kegg_pathway_126
BoNT Light Chain Types B, D, and F cleave VAMP/Synaptobrevin	0.024	reactome_pathway_647
Amyloids	0.026	reactome_pathway_643
Nucleotide-binding domain, leucine rich repeat containing receptor (NLR) signaling pathways	0.026	reactome_pathway_465
BH3-only proteins associate with and inactivate anti-apoptotic BCL-2 members	0.034	reactome_pathway_517
Axonal growth inhibition (RHOA activation)	0.039	reactome_pathway_984
Vitamin C (ascorbate) metabolism	0.039	reactome_pathway_188
Import of palmitoyl-CoA into the mitochondrial matrix	0.039	reactome_pathway_70
Downregulation of ERBB4 signaling	0.039	reactome_pathway_1023
p75NTR regulates axonogenesis	0.043	reactome_pathway_982
Caspase-8 is formed from procaspase-8	0.043	reactome_pathway_507
Activation of Pro-Caspase 8	0.043	reactome_pathway_508
Endosomal/Vacuolar pathway	0.043	reactome_pathway_413
Phase II conjugation	0.047	reactome_pathway_249
SLBP independent Processing of Histone Pre-mRNAs	0.048	reactome_pathway_1111
Arachidonic acid metabolism	0.050	kegg_pathway_56

Pathways in italic are also found in ingenuity pathway analysis.

Table 3-5 : Replication of causality genes in the bronchial airway epithelial and the regional lung tissue datasets.

	Replication datasets	
	Bronchial airway epithelium[13]	Regional lung tissue[14]
FEV_1 % predicted		
NCR3	True	False
CST3	True	
CSTA	True*	
PPARGC1A	False	True
FLCN	False	
BCL2L1	True	False
CADM2	False	
TNFRSF10B	True*	
FEV_1/FVC		
GSTO2	False	
DEPDC6	False	
CD22	True	True*
MUC22		
SPINK5	True	

*Significant in the original studies[13,14].

Figures

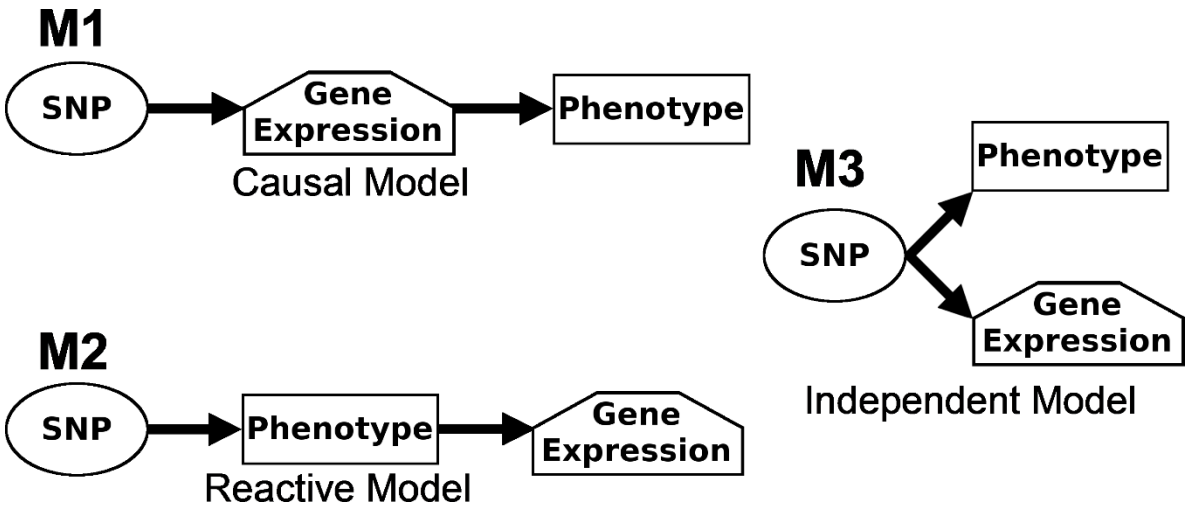

Figure 3-1 : Causality models showing the relationships between the expression of a gene, a phenotype, and a SNP.

Three models are depicted including a Causal Model (M1), a Reactive Model (M2), and an Independent Model (M3). Model 1 (M1) is the simplest, and states that the genotype at an eQTL-SNP acts directly on gene expression pattern to produce the phenotype. Model 2 (M2) states that the gene expression pattern is reactive to the phenotype, and Model 3 (M3) states that the gene expression pattern and phenotype are independent.

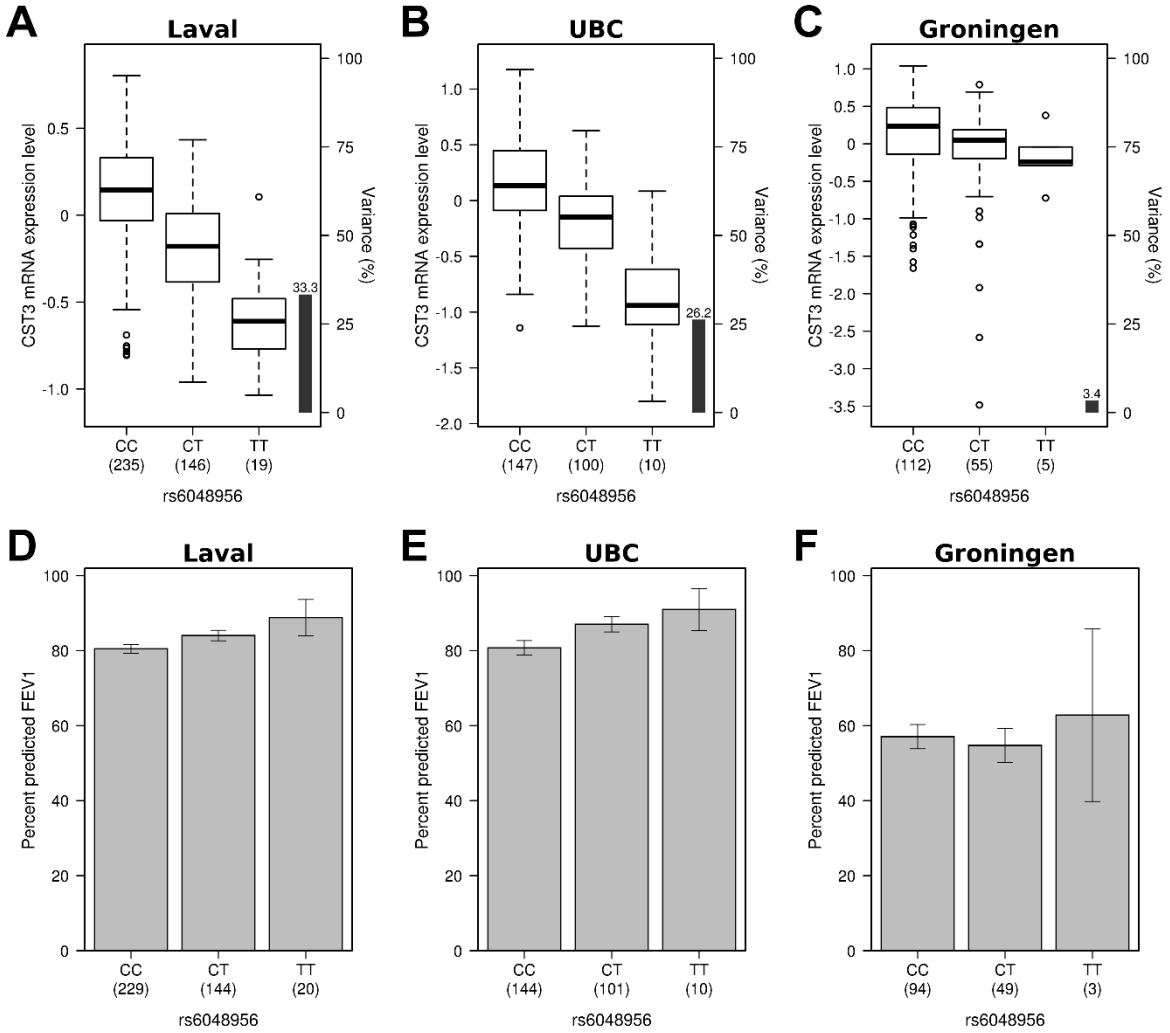

Figure 3-2 : Direction of effects for causality model with rs6048956, mRNA expression of CST3, and FEV1 % predicted.

The left, center, and right panels show the results for Laval, UBC and Groningen samples, respectively. A), B), and C) are boxplots of gene expression levels in the lung for CST3 according to genotype groups for SNP rs6048956. The left y-axis shows the mRNA expression levels for CST3. The x-axis represents the three genotype groups for SNP rs6048956. The right y-axis shows the proportion of the gene expression variance explained by the SNP (black bar). Box boundaries, whiskers, and center mark in boxplots represent the first and third quartiles, the most extreme data point which is no more than 1.5 times the interquartile range, and median, respectively. D), E), and F) are barplots of the mean and standard error of FEV1 % predicted according to genotype groups for SNP rs6048956.

Online Data Supplement

Methods

Subject selection

The methods for subject selection and phenotyping as well as for interrogation of gene expression and genotype were recently described[1]. The lung tissue used for discovery of eQTLs was from 1,111 human subjects who underwent lung surgery at three academic sites, Laval University, University of British Columbia (UBC), and University of Groningen, henceforth referred to as Laval, UBC, and Groningen, respectively. For the present study, the principal aim was to examine smoking-related airway obstruction. Thus, we excluded subjects whose lung function may have been influenced by lung diseases other than COPD. Exclusion criteria were: 1) missing data for lung function, i.e. for both Forced Expiratory Volume in 1 second as a percentage of its predicted value (FEV_1 % predicted) and that divided by Forced Vital Capacity (FEV_1/FVC) and 2) patients with asthma, extensive pneumonia, cystic fibrosis, bronchiectasis, pulmonary fibrosis, pulmonary hypertension, primary bullous emphysema, mesothelioma, diffuse alveolar damage and alpha-1-antitrypsin deficiency. We did not exclude individuals with lung cancer, focal pneumonia or atelectasis, but in these cases lung tissue samples were taken as far away as possible from involved areas. Although lung function in theory could be compromised in these cases, the vast majority of subjects included in the biobanks from the three participating sites had relatively small, non-obstructing lung tumors, which were unlikely to significantly interfere with the subjects' lung function. Of the 1,111 subjects in whom there were data on genotype and gene expression, 848 with sufficient phenotypic information and without a lung disease (other than COPD and lung cancer) were included in the analysis. The demographic and clinical features of the subjects in the three cohorts are described in **Table 3-1**.

COPD phenotypes

For the primary analyses, three phenotypes were used: FEV_1 % predicted and FEV_1/FVC as continuous variables, and COPD defined dichotomously based on an $FEV_1/FVC < 0.7$ cutoff. We used post-bronchodilator spirometry when available; otherwise, pre-bronchodilator values were used.

Genome-wide association study

Genome-wide association was performed on the three phenotypes using linear or logistic regression models. Single-marker association tests were run within each cohort adjusting for age, gender and smoking status. Furthermore, we conducted a fixed-effects meta-analysis combining the three cohorts using inverse standard error weighting. We performed genomic control (GC) correction for individual cohorts and for the meta-analysis. To avoid over-correction, we computed the genomic inflation factor (□) on the 90% of SNPs with the largest p-values. The □ estimates were small (≤1.03) for the individual cohort analyses and the meta-analysis.

Causality Models

We evaluated three competing causality models to describe the relationship between lung eQTL-SNPs, RNA expression and COPD phenotypes (**Figure 3-1**). We tested three linear equations in describing the molecular relationships,

$$T_i = \alpha_1 + \beta_1 L_i + \varepsilon_{1i} \qquad \text{(eq 1)}$$

$$G_i = \alpha_2 + \beta_2 T_i + \beta_3 L_i + \varepsilon_{2i} \qquad \text{(eq 2)}$$

$$T_i = \alpha_3 + \beta_4 G_i + \beta_5 L_i + \varepsilon_{3i} \qquad \text{(eq 3)}$$

A SNP at a specific locus is denoted by L, gene expression for a specific transcript by G, and a measured clinical endpoint by T (i.e. FEV_1 % predicted, FEV_1/FVC or COPD yes/no). Dependencies are likely to exist between certain pairs of covariates in the preceding three models. We inferred a causal relationship using conditional correlation[2]. We required a p-value for $\beta_1 < 1 \times 10^{-3}$ before examining a particular (L, G, T) variable

triplet. If a p-value for $\beta_3 < 0.05$ and p-value for $\beta_5 > 0.05$, the causal model was selected. If the converse was true, a reactive model was selected. If both p-values were < 0.05, an independence model was selected. If both p-values were > 0.05 then no call was made. Model 1 indicates a potential causal relationship and was the main interest of this work.

Analyses were performed in the three cohorts separately and then combined into a meta-analysis. We then conducted sample bootstrapping and repeated the causality test for N=1000 realizations. The reliability score of the molecular relationship is the fraction of bootstrap realizations that supports the call observed.

Manual curation of significant and reliable causal models

To explore the biology and possible role of genes in the causal model, experts in the pathobiology of COPD at each of the participating sites manually reviewed the literature for functionality and known genetic associations. Several strategies were used including PubMed searches using the gene's name and COPD or emphysema, interrogation of NCBI PubMed Gene site, GeneRIFs (Gene Reference Into Functions database), PheGenI for eQTL, and phenotype association data.

Pathway analyses on causality results

Pathway analyses were performed on causality results in order to search for enrichments of specific gene pathways. Genes that fitted Model 1 (causal) with reliability scores > 0.8 were analyzed using the Ingenuity Pathways Analysis (IPA, Ingenuity® Systems, www.ingenuity com). Causal genes were mapped to corresponding gene objects in IPA using official gene symbols and overlaid onto canonical pathways contained in the Ingenuity Pathways Knowledge Base. The latter analysis identified pathways from the IPA library of canonical pathways that were most enriched with causal genes. The significance of the association with canonical pathways was determined using a right-tailed Fisher's exact test. This test compared the number of causal genes versus total genes in a canonical pathway beyond that expected by chance alone. A total of 181 canonical pathways were tested. A nominal p-value < 0.05 was considered significant. Similarly the same list of causality genes was examined for enrichment in gene ontology and functional categories

using MetaCore™ (version 6.12, build 42289, GeneGo, Inc.) and mapped to both the Gene Ontology (GO) project and proprietary ontologies in the MetaCore™ knowledge database. The significance of functional enrichment of genes was determined by using a False Discovery Rate (FDR) cutoff of 0.05. Network analysis was also carried out using canonical networks in the Metacore™ knowledge database. Direct interaction networks based on our gene lists were built using seed nodes and their direct interactions were assessed using curated, known interactions. Pathway analysis was also carried out using Partek Genomic Suite 6.6 software (Partek GS, version 6.12.0907, www.partek.com). Causality genes were mapped to official gene symbols in Partek GS and parsed to Partek Pathway software. Partek Pathway interrogates the REACTOME (www.reactome.org) and KEGG (www.genome.jp/kegg/) databases and identifies canonical pathways that are enriched with causality genes. Enrichment score p-values were derived using Fisher's Exact test. A p-value < 0.05 was considered significant.

Replication studies

Bronchial airway epithelial dataset. Bronchial airway brushings were obtained during bronchoscopy from active and former smokers who were enrolled in a lung cancer screening program at the British Columbia Cancer Research Agency[3]. Institutional Review Board approval was obtained at participating institutions, and all subjects provided written informed consent. RNA isolated from bronchial brushings of 238 lung cancer-free active and former smokers with and without COPD was profiled using Affymetrix Human Gene 1.0 ST Arrays (Affymetrix Inc., Santa Clara, CA, USA). Microarray data were deposited in the Gene Expression Omnibus (GSE37147). Gene expression estimates were derived and normalized as previously described[3]. Gene expression levels associated with FEV_1 % predicted, FEV_1/FVC and the presence of COPD were determined using linear modeling as previously described[3].

Regional lung tissue dataset. Whole lungs were explanted from patients with severe COPD (n=6) and from donors (n=2). Each lung was sampled in eight consecutive regions from the apex to the base of the lung. This study was approved by the institutional review board and written informed consent was obtained from each patient prior to surgery or from the next of kin of the persons who the donated lung. The gene expression profiles were

obtained in 64 samples (8 patients x 8 regions) using the Human Exon 1.0 ST array. The gene expression dataset was deposited in the Gene Expression Omnibus (GSE27597). Details of sample collection and processing as well as analytical methods to normalize and obtain gene expression values have been previously described[4]. The degree of emphysema in each sample was quantified by measuring the mean linear intercept (Lm) on micro-computed tomography (CT) scans for tissues. Lm represents a morphological measurement of alveolar destruction and is a surrogate for emphysema severity. In this study, genes were associated with regional emphysema severity within the same lung using linear models as described before[4].

Tables

Table 3-S1 : List of causal models with a confidence score >0.8.

See the attached file.

Table 3-S2 : Models of biological relevance identified by manual curation

SNPs	Gene (probe set)	Gene role & references
FEV1 % predicted		
rs769178	*NCR3* (100125842_TGI_at)	SNPs near *NCR3* gene were associated with lung function[5].
rs6048956	*CST3* (100307577_TGI_at)	Protease-antiprotease balance[6].
rs6515375	*CST3* (100125967_TGI_at)	Protease-antiprotease balance[6].
rs2270859	*CSTA* (100148334_TGI_at)	Expression of CSTA is modulated by genotype, smoking, COPD and lung cancer[7].
rs4550905	*PPARGC1A* (100131093_TGI_at)	The PPAR□-PGC-1α pathway regulates antioxidant genes and protects against COPD in rats[8].
rs3803761	*FLCN* (100135396_TGI_at)	Mutations in *FLCN* cause Birt-Hogg-Dubé syndrome, a monogenic disorder characterized by spontaneous pneumothorax. Genetic variants in *FLCN* were not associated with severe, early-onset COPD[9].
rs1543438	*BCL2L1* (100158784_TGI_at)	Cigarette smoke extract induces the expression of BCL2L1in human dendritic cells and augments survival of these cells in COPD patients[10]
rs9880397	*CADM2* (100162763_TGI_at)	SNPs in *CADM2* were associated with lung function and asthma (unpublished).
rs2466183	TNFRSF10B	Involved in T cell and eosinophil regulation in

	(100153254_TGI_at)	bronchial smooth muscle cell death in asthma[11]. Increased expression of the protein in the lung of subjects with emphysema[12].
FEV_1/FVC		
rs17754977	*GSTO2* (100132911_TGI_at)	GSTO enzymes protect against oxidative stress. Involved in the biotransformation of arsenic found in cigarette smoke[13]. SNPs in *GSTO2* were associated with FEV_1, FVC[14] and COPD[15].
rs9987135	*DEPDC6* (100154484_TGI_at)	DEPDC6 is a negative regulator of mTORC1 and mTORC2 signaling pathways. Decreased TOR activity has been found to slow aging in yeast, worms, flies, and mice[16]. The mTOR inhibitor, rapamycin, increases the lifespan in mice and reduces bronchial hyperresponsiveness and airway remodelling[17].
rs10411704	*CD22* (100154732_TGI_at)	CD22 is present on the surface of (mature) B cells and prevents over activation of the immune system and development of autoimmune system.
rs12179536	*MUC22* (100304000_TGI_at)	Polymorphisms in *MUC22* were associated with panbronchiolitis[18]. MUC22 is expressed in the lung.
rs2287765	*SPINK5* (100305138_TGI_at)	*SPINK5* is a candidate gene for asthma and allergy[19]. Upregulation of SPINK5 in epithelial cell line increases inflammatory responses[20].

Figures

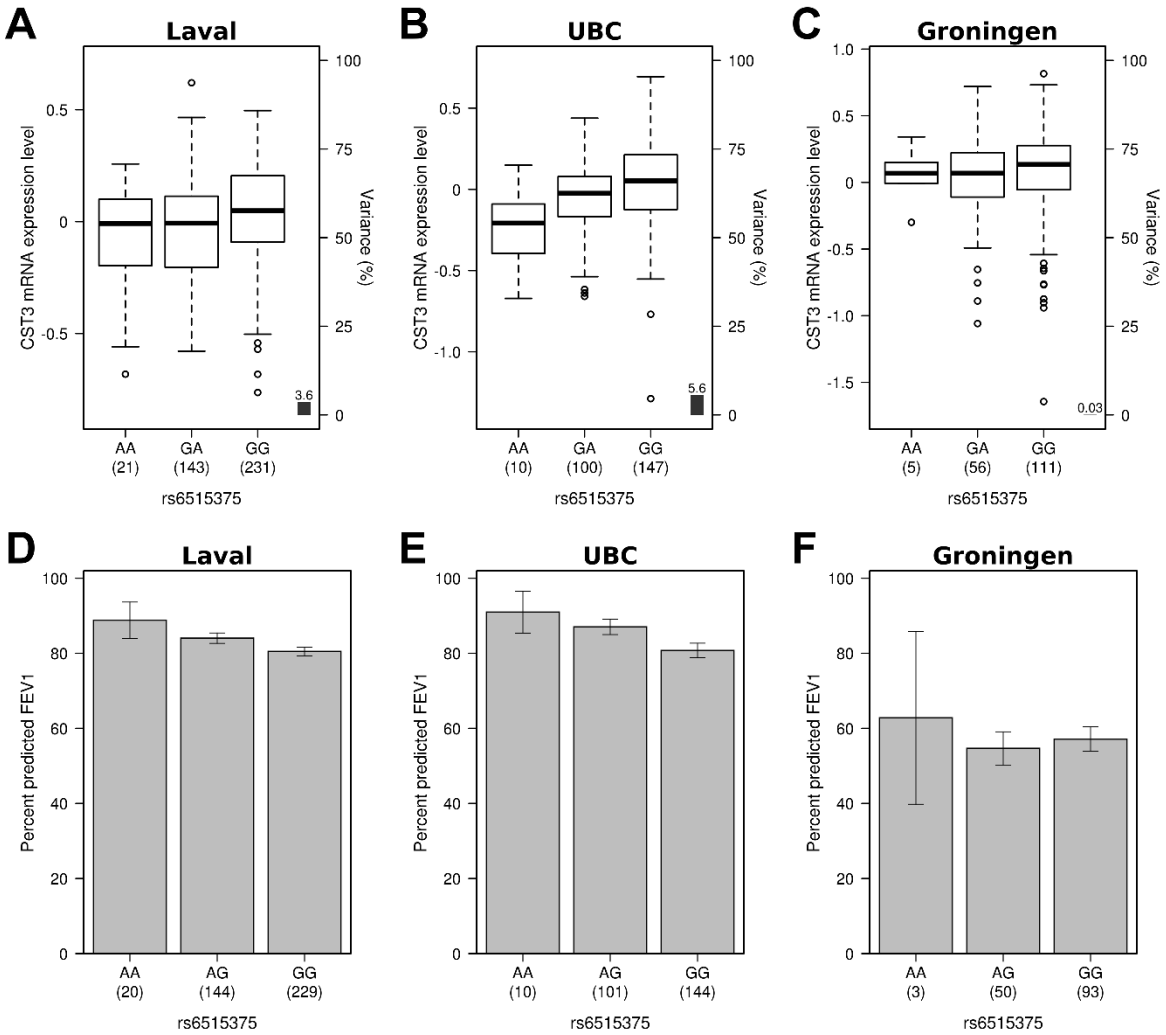

Figure 3-S1 : Direction of effects for causality model with rs6515375, mRNA expression of CST3, and FEV1 % predicted. Data are presented as described in Figure 3-2.

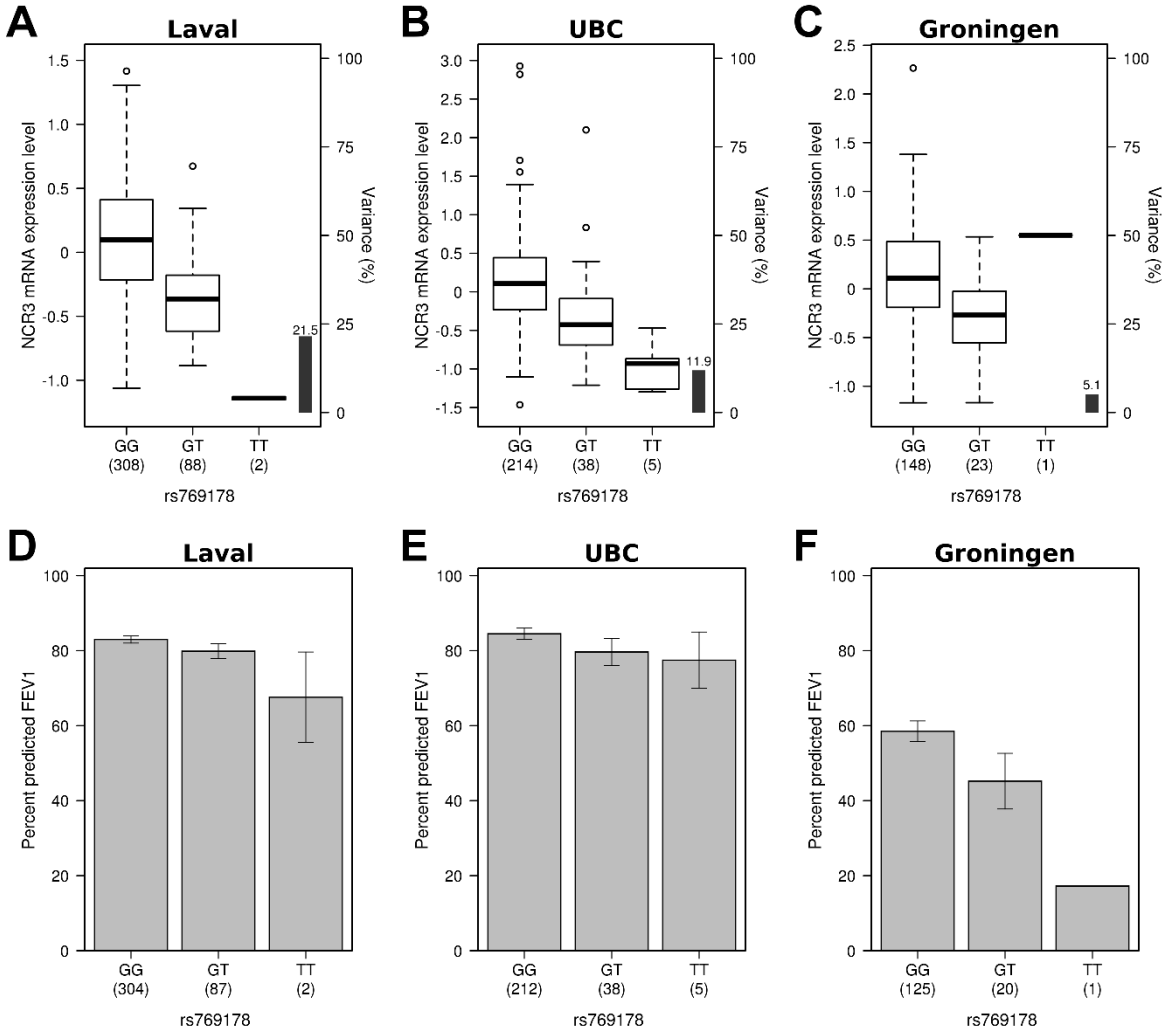

Figure 3-S2 : Direction of effects for causality model with rs769178, mRNA expression of NCR3, and FEV1 % predicted. Data are presented as described in Figure 3-2.

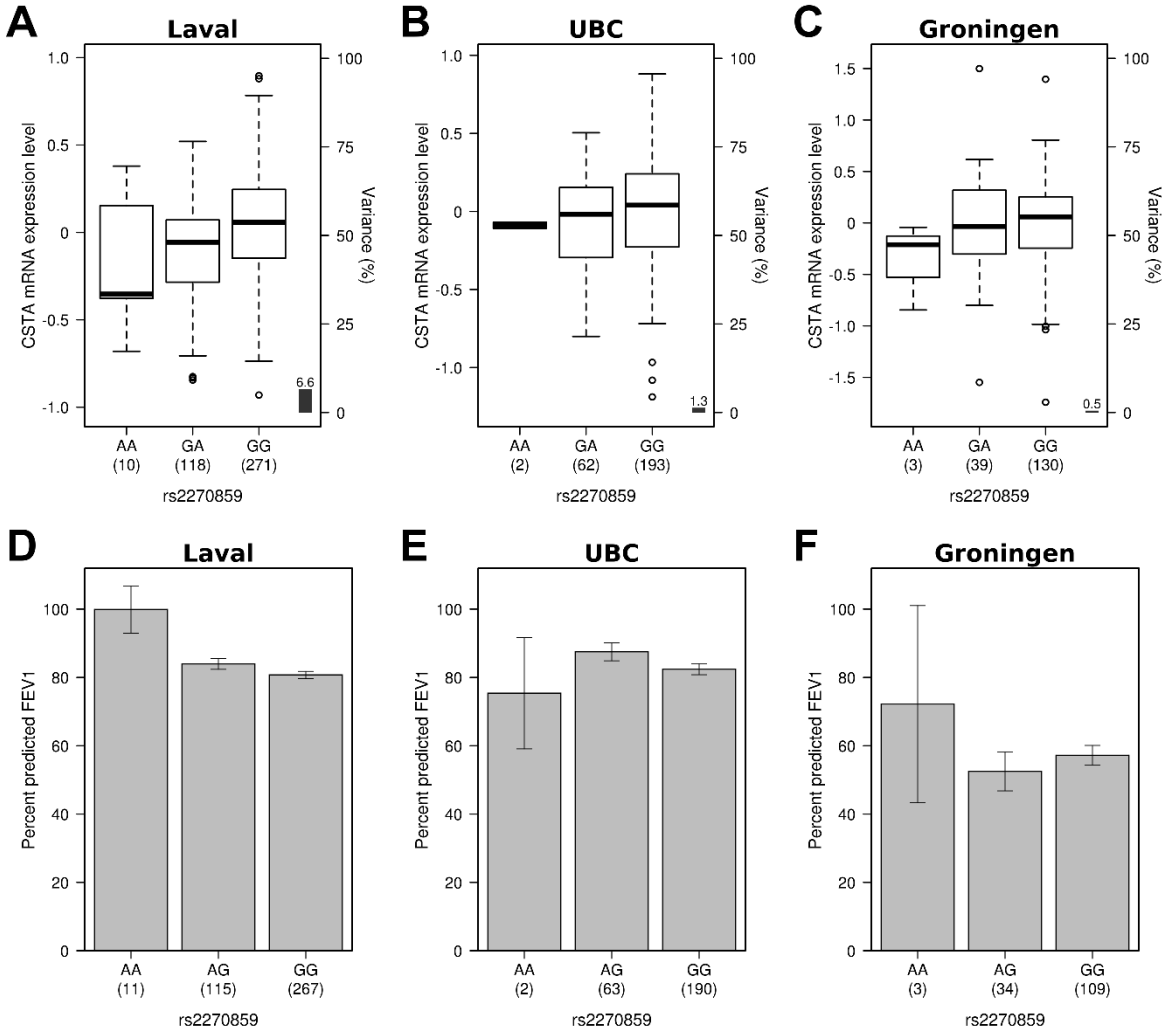

Figure 3-S3 : Direction of effects for causality model with rs2270859, mRNA expression of CSTA, and FEV1 % predicted. Data are presented as described in Figure 3-2.

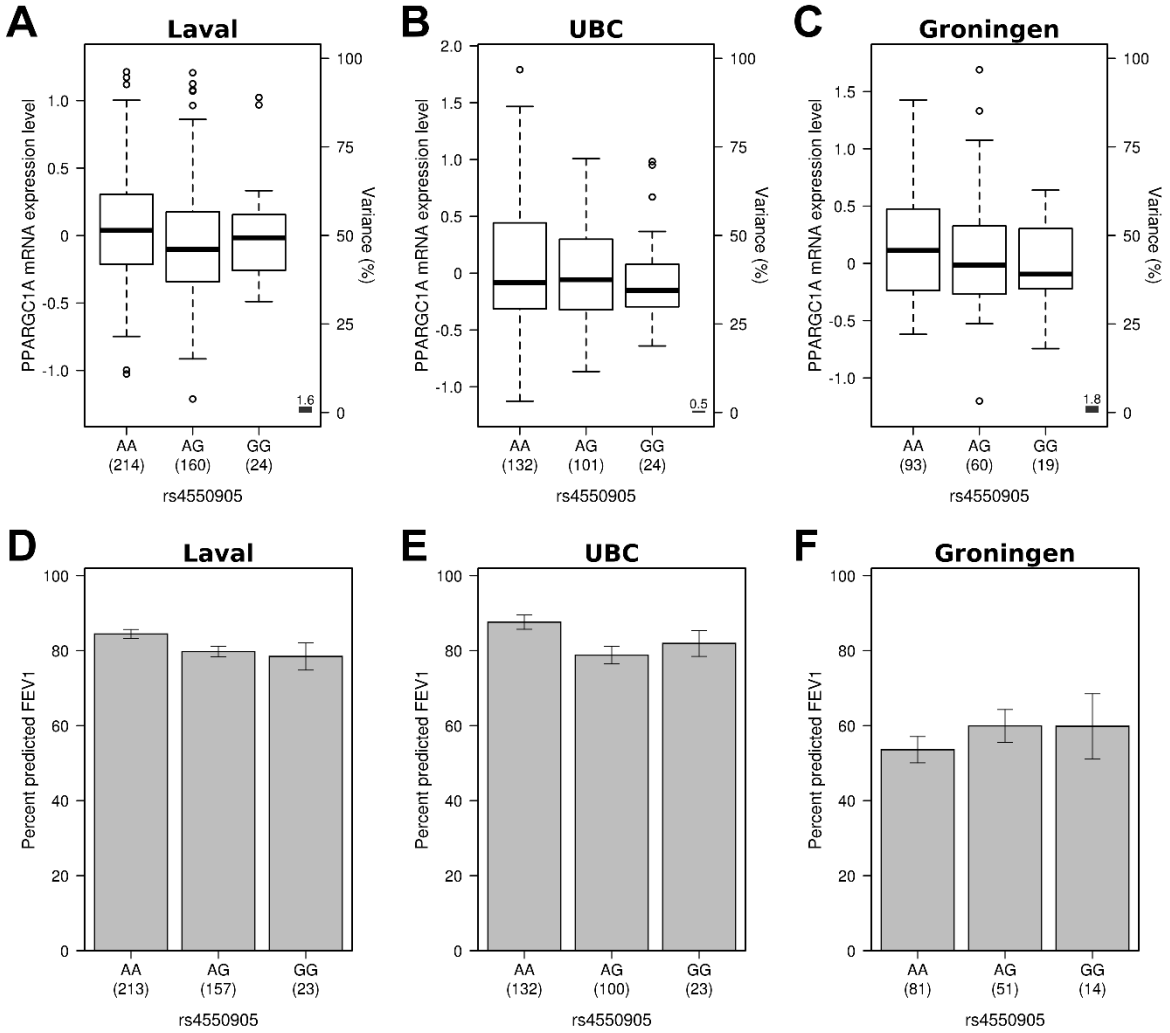

Figure 3-S4 : Direction of effects for causality model with rs4550905, mRNA expression of PPARGC1A, and FEV1 % predicted. Data are presented as described in Figure 3-2.

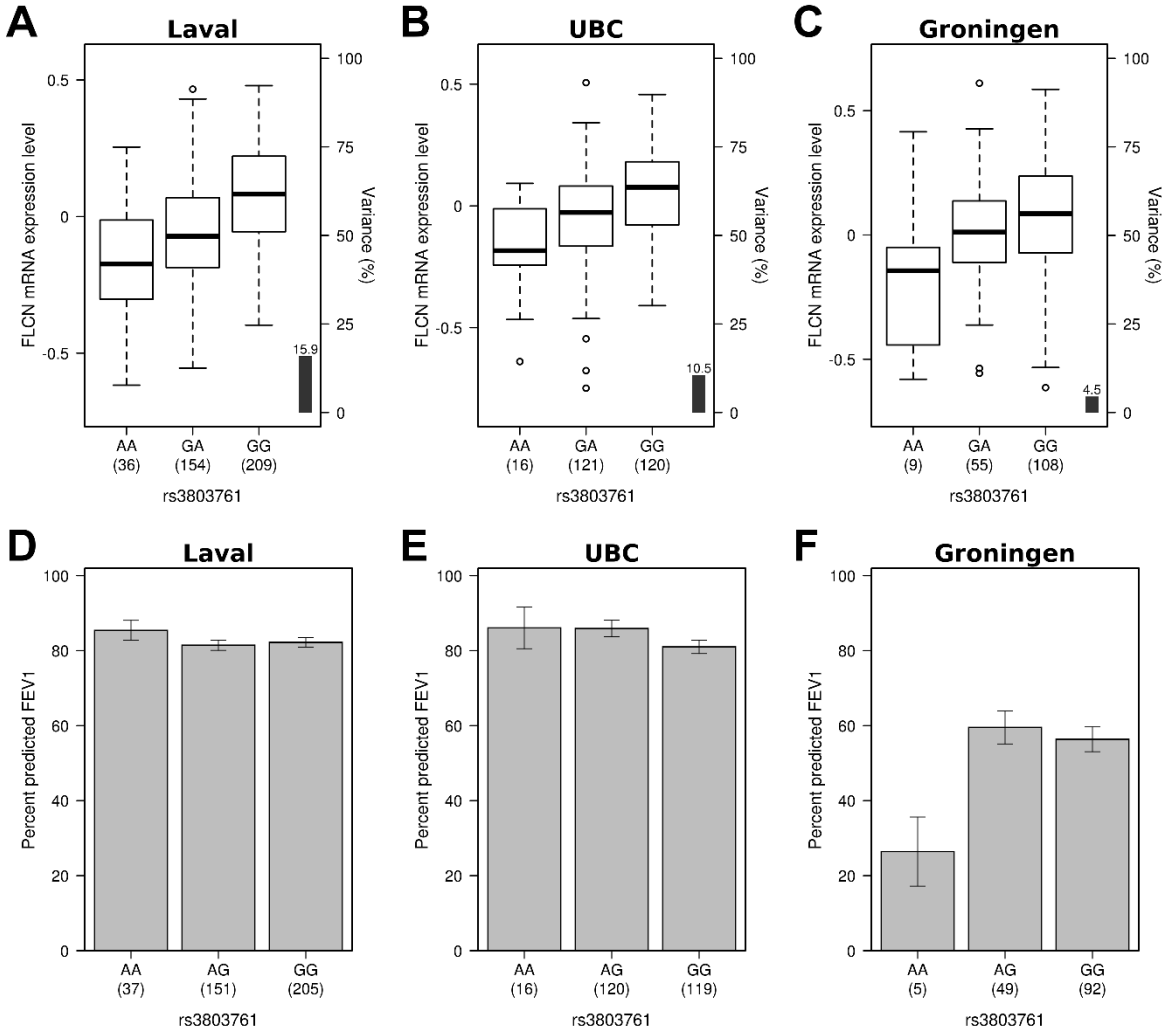

Figure 3-S5 : Direction of effects for causality model with rs3803761, mRNA expression of FLCN, and FEV1 % predicted. Data are presented as described in Figure 3-2.

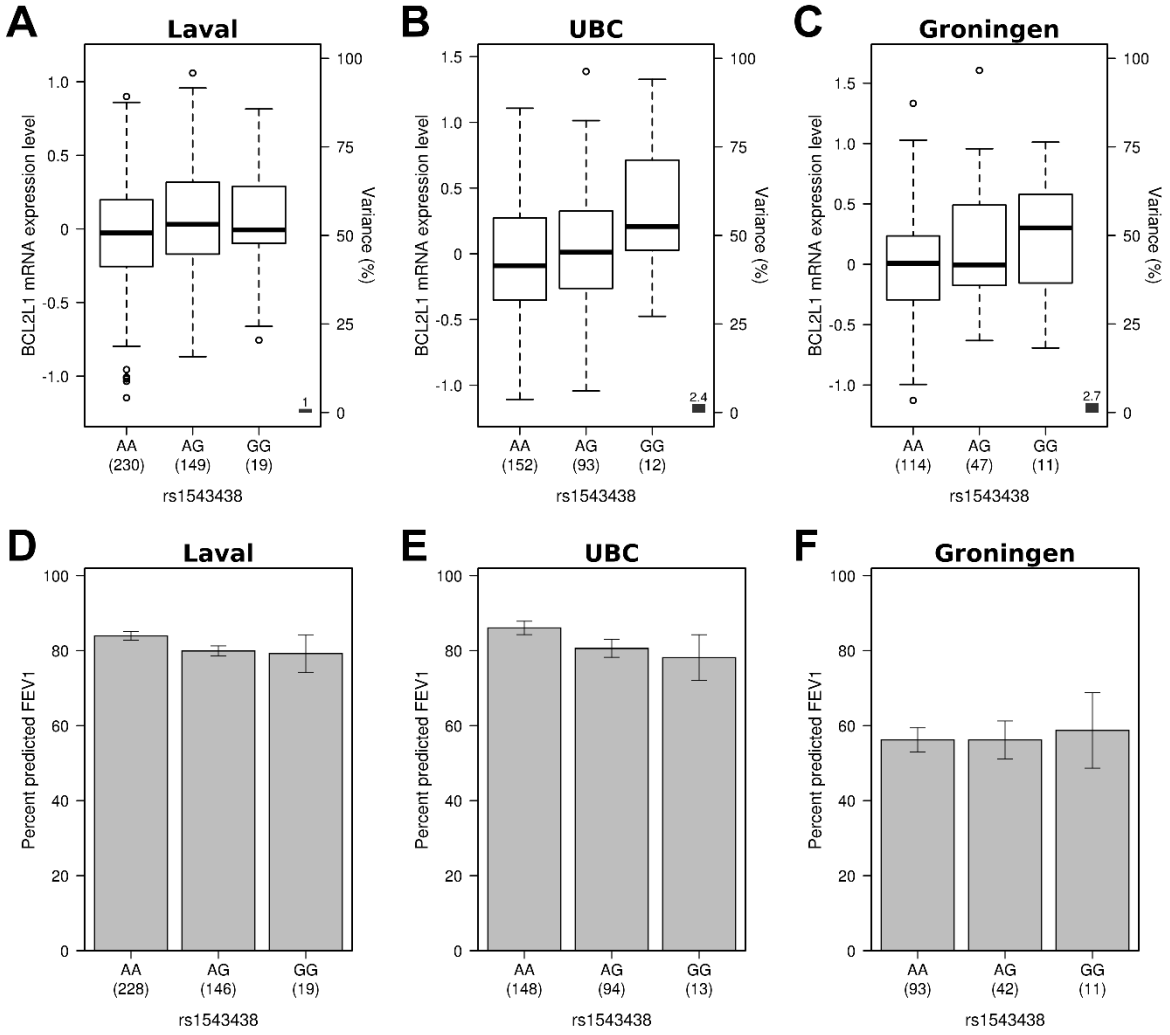

Figure 3-S6 : Direction of effects for causality model with rs1543438, mRNA expression of BCL2L1, and FEV1 % predicted. Data are presented as described in Figure 3-2.

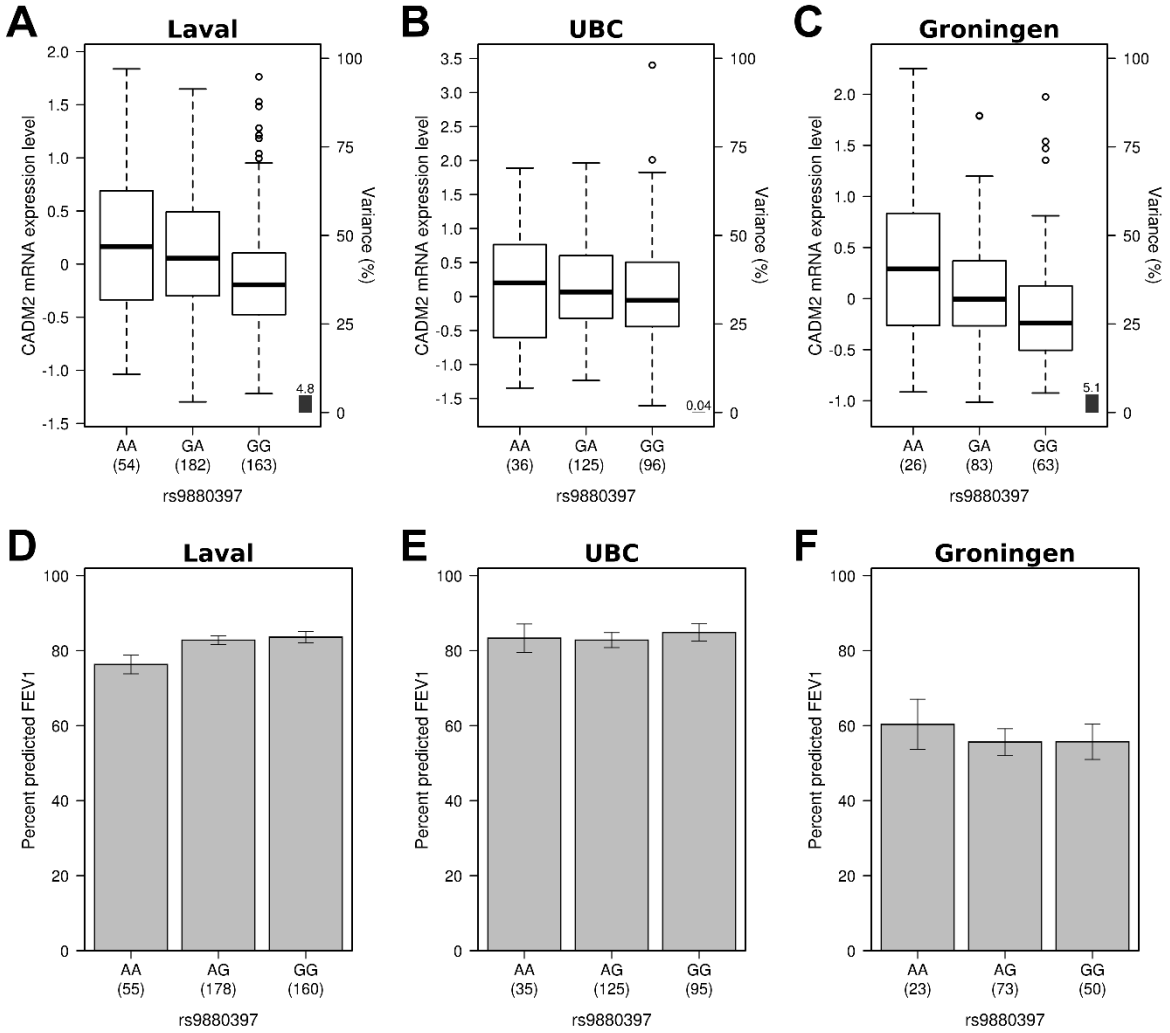

Figure 3-S7 : Direction of effects for causality model with rs9880397, mRNA expression of CADM2, and FEV1 % predicted. Data are presented as described in Figure 3-2.

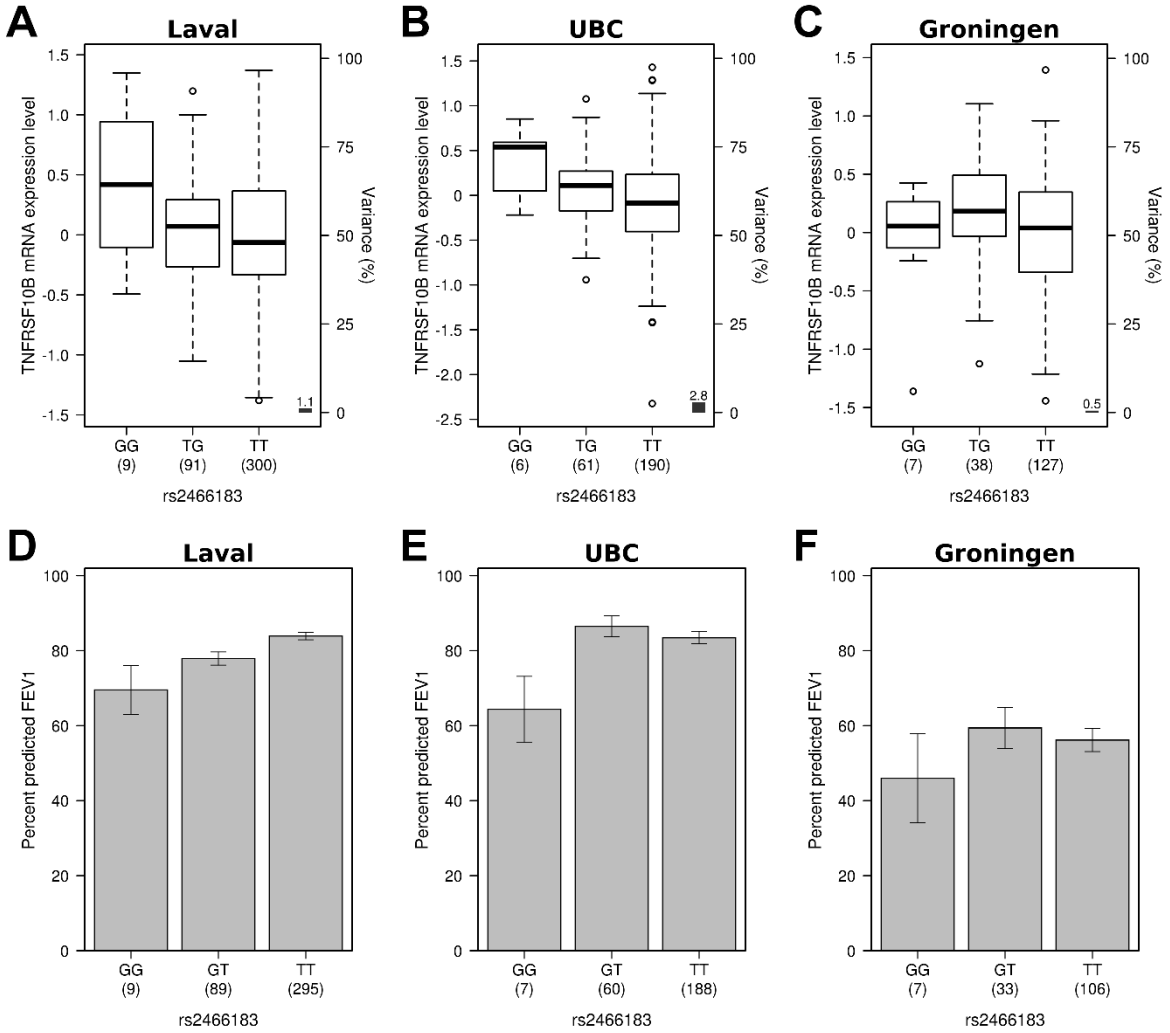

Figure 3-S8 : Direction of effects for causality model with rs2466183, mRNA expression of TNFRSF10B, and FEV1 % predicted. Data are presented as described in Figure 3-2.

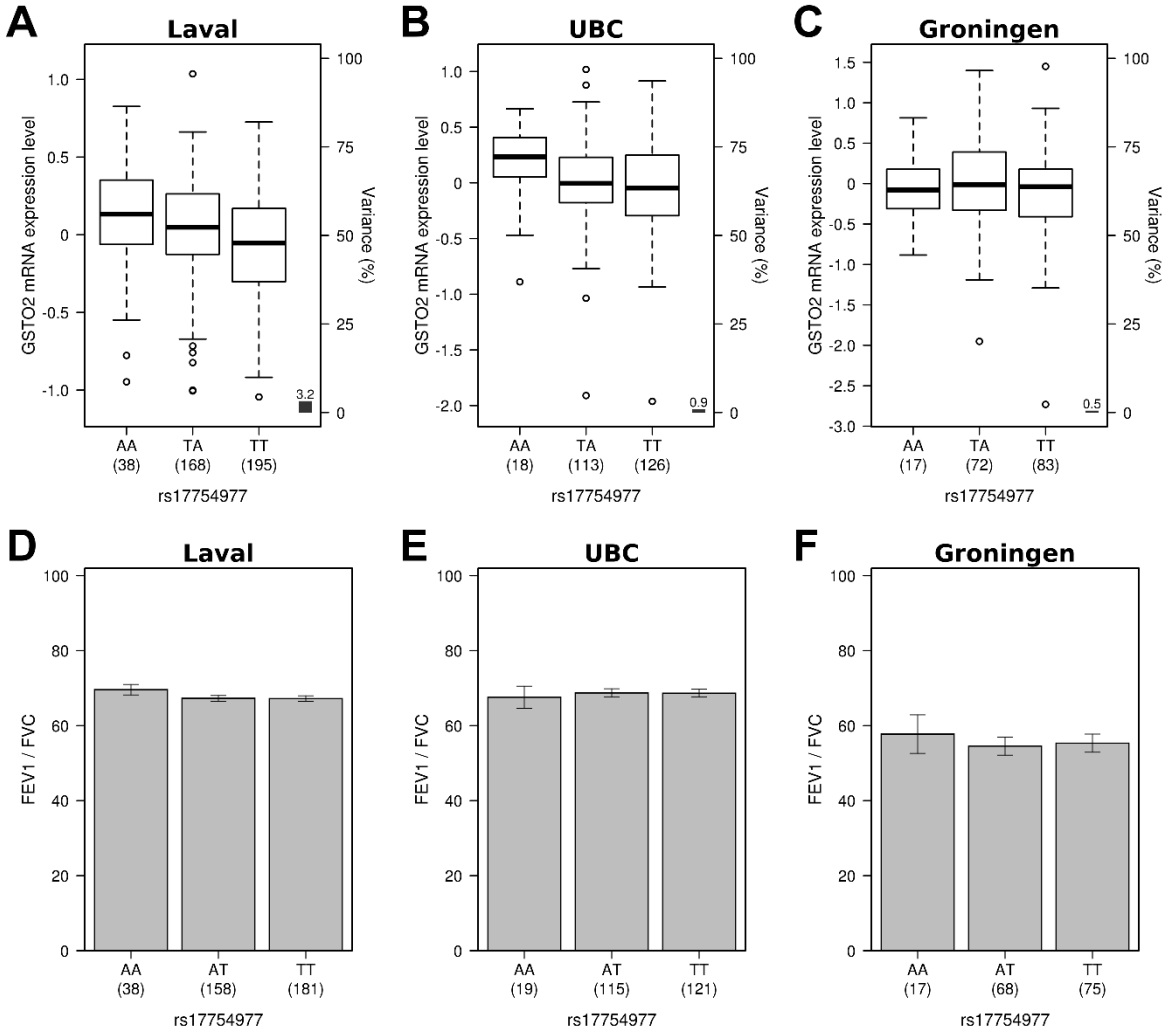

Figure 3-S9 : Direction of effects for causality model with rs17754977, mRNA expression of GSTO2, and FEV1/FVC. Data are presented as described in Figure 3-2.

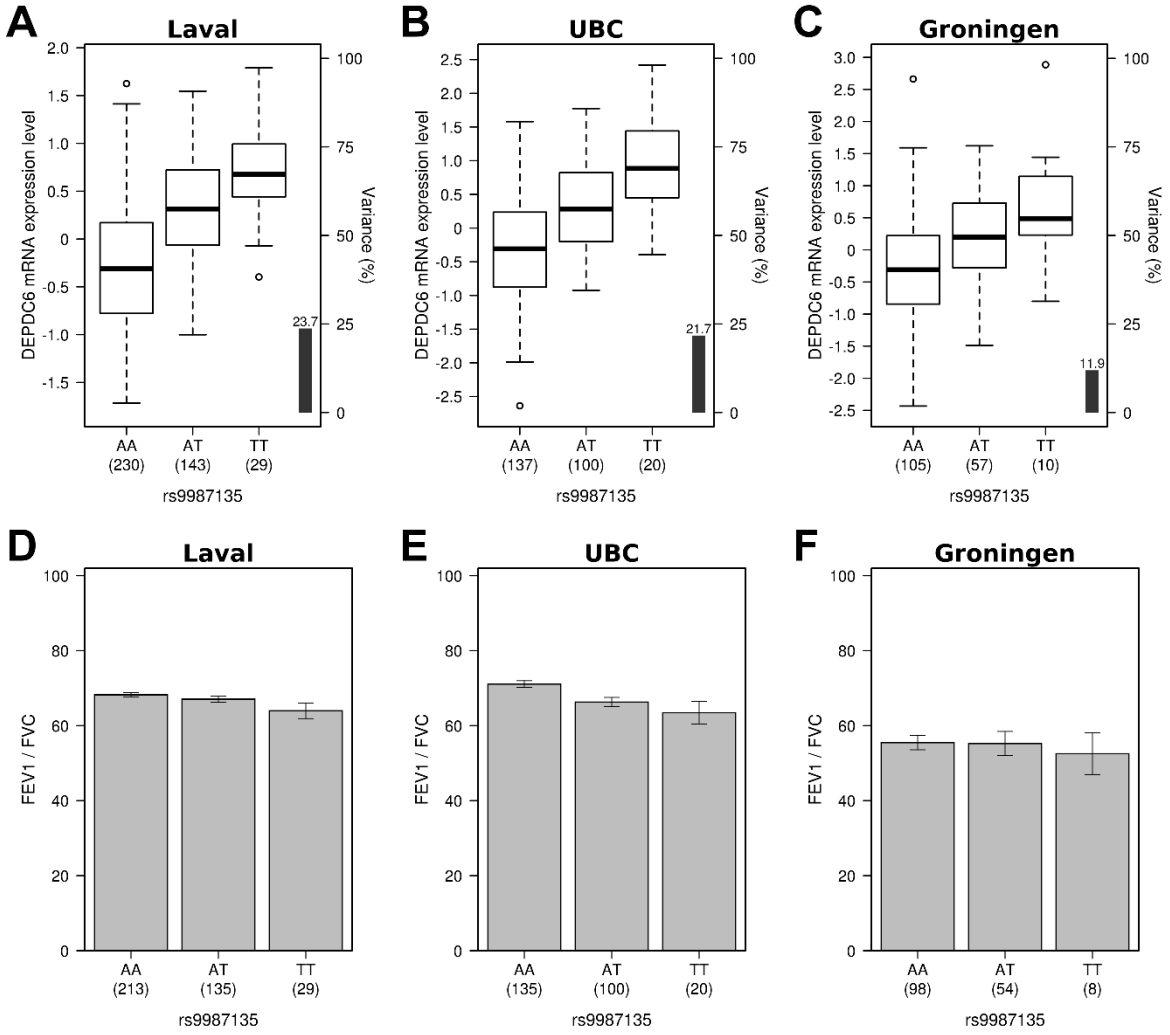

Figure 3-S10 : Direction of effects for causality model with rs9987135, mRNA expression of DEPDC6, and FEV1/FVC. Data are presented as described in Figure 3-2.

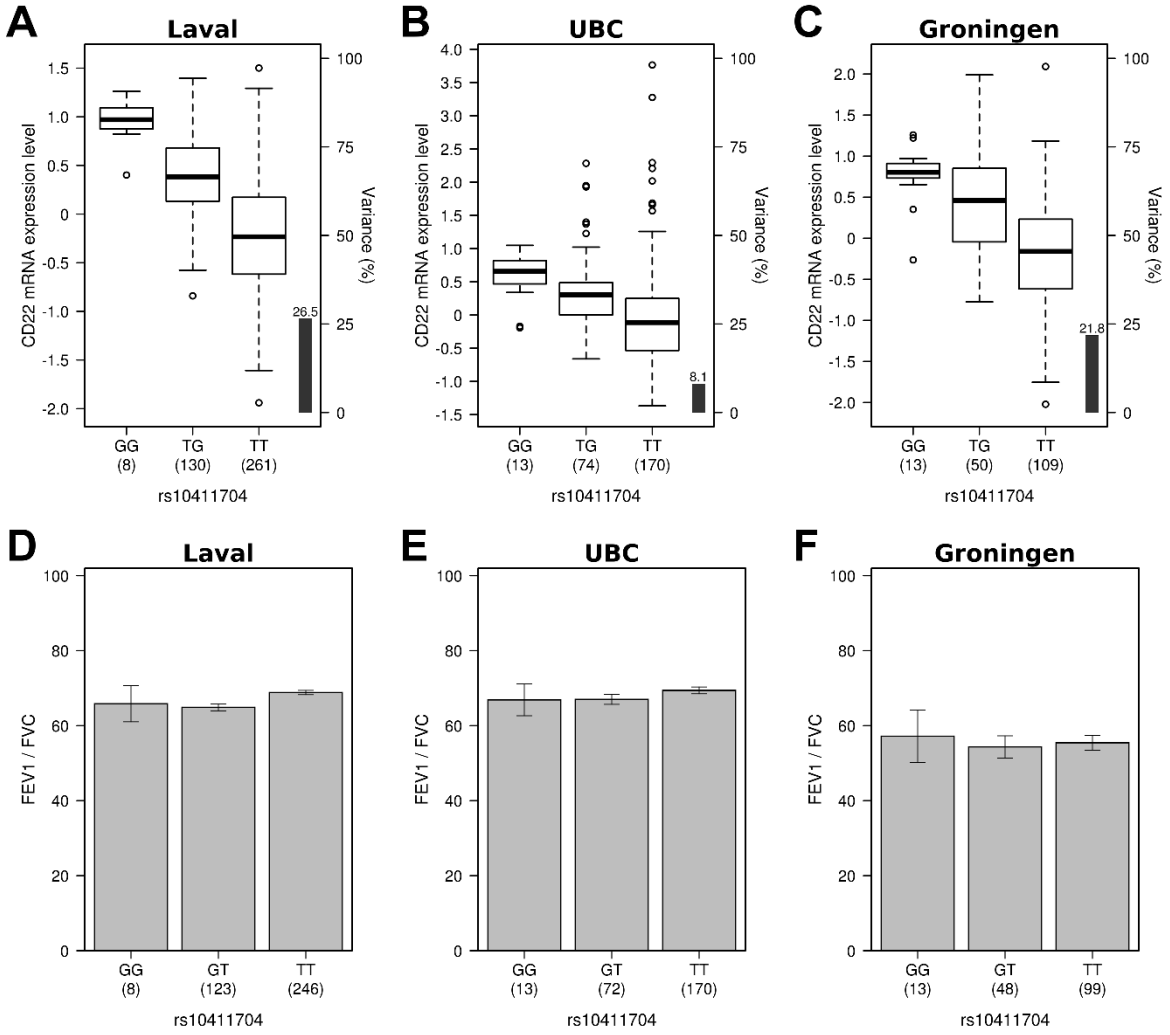

Figure 3-S11 : Direction of effects for causality model with rs10411704, mRNA expression of CD22, and FEV1/FVC. Data are presented as described in Figure 3-2.

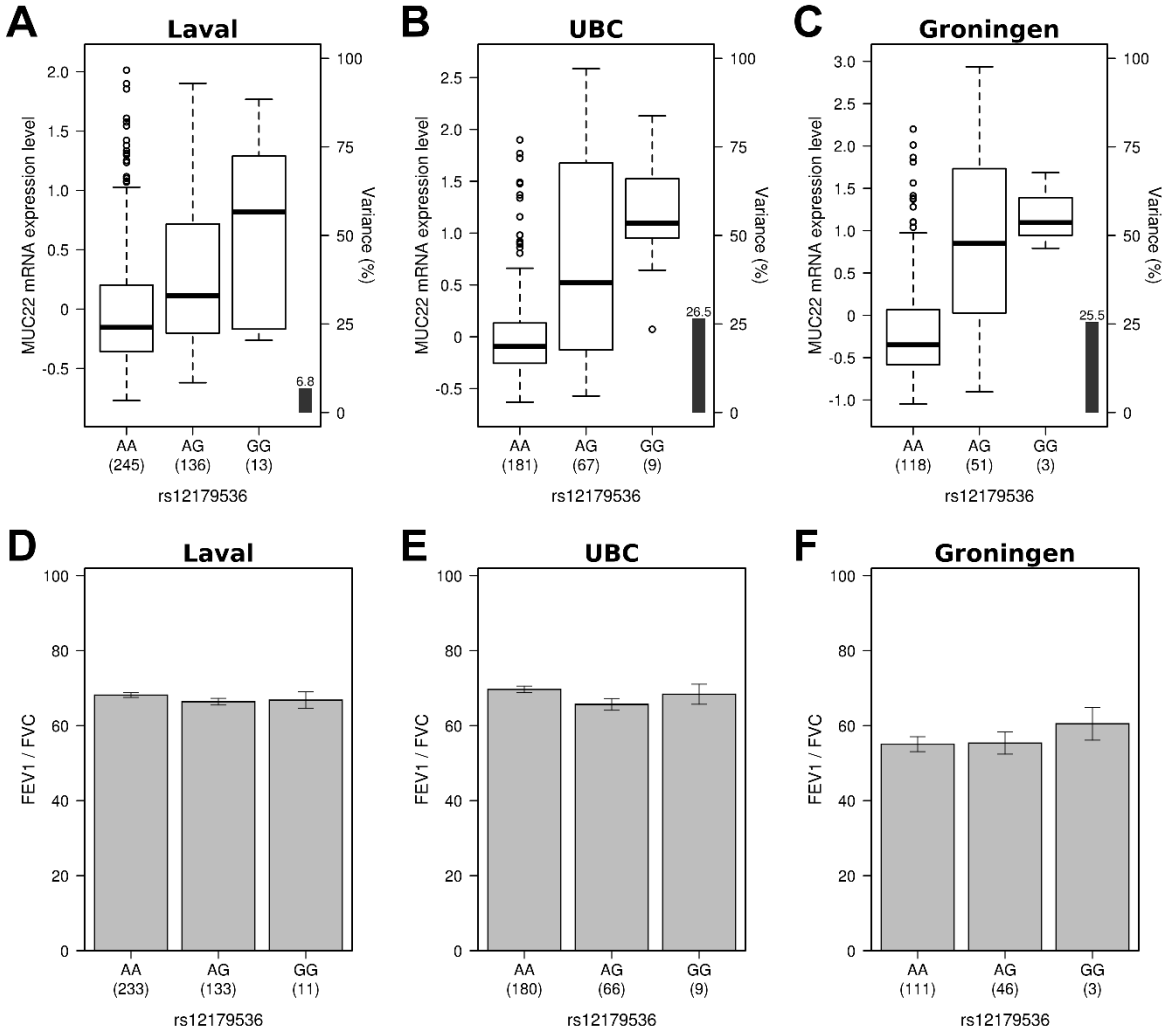

Figure 3-S12 : Direction of effects for causality model with rs12179536, mRNA expression of MUC22, and FEV1/FVC. Data are presented as described in Figure 3-2.

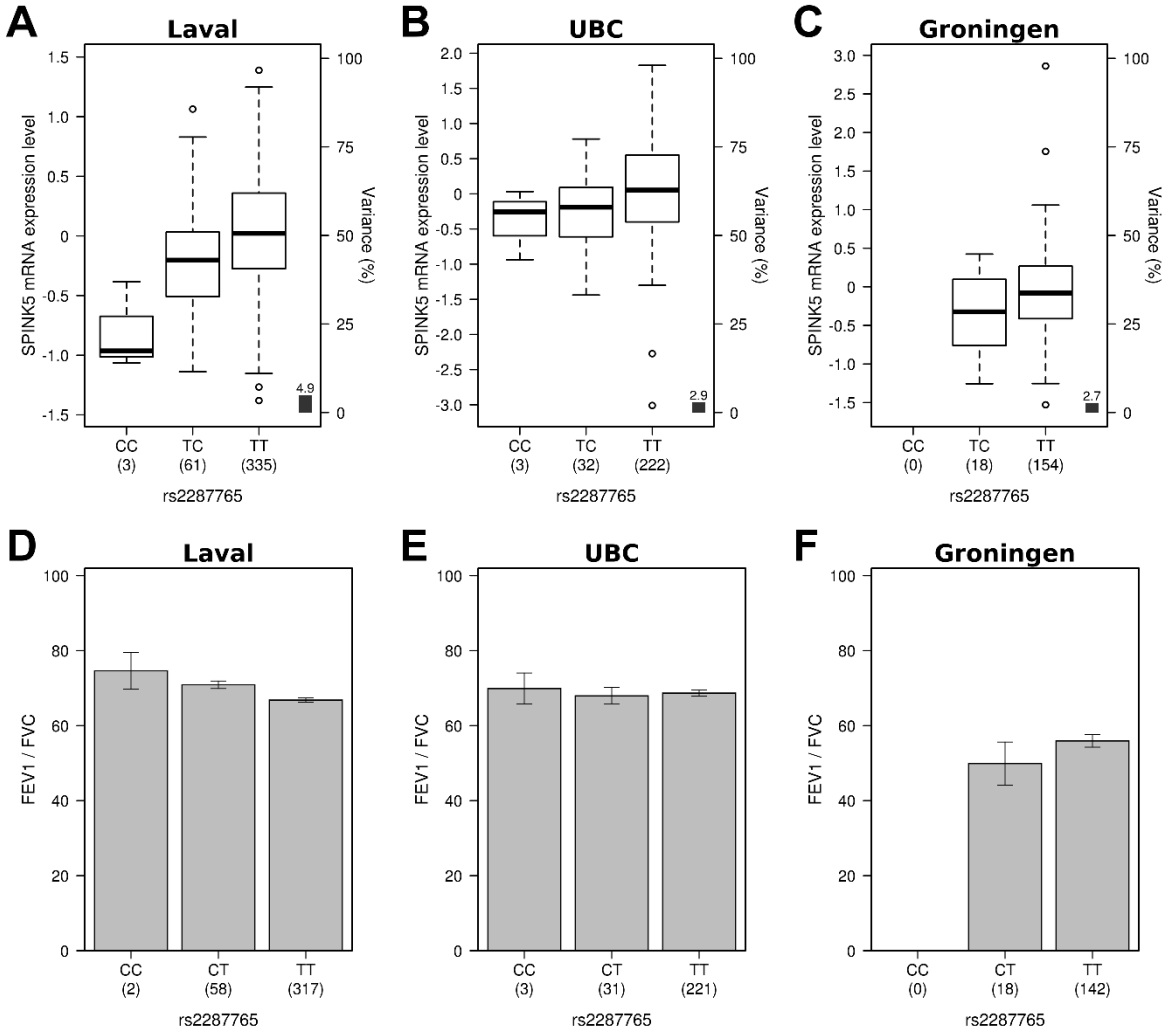

Figure 3-S13 : Direction of effects for causality model with rs2287765, mRNA expression of SPINK5, and FEV1/FVC. Data are presented as described in Figure 3-2.

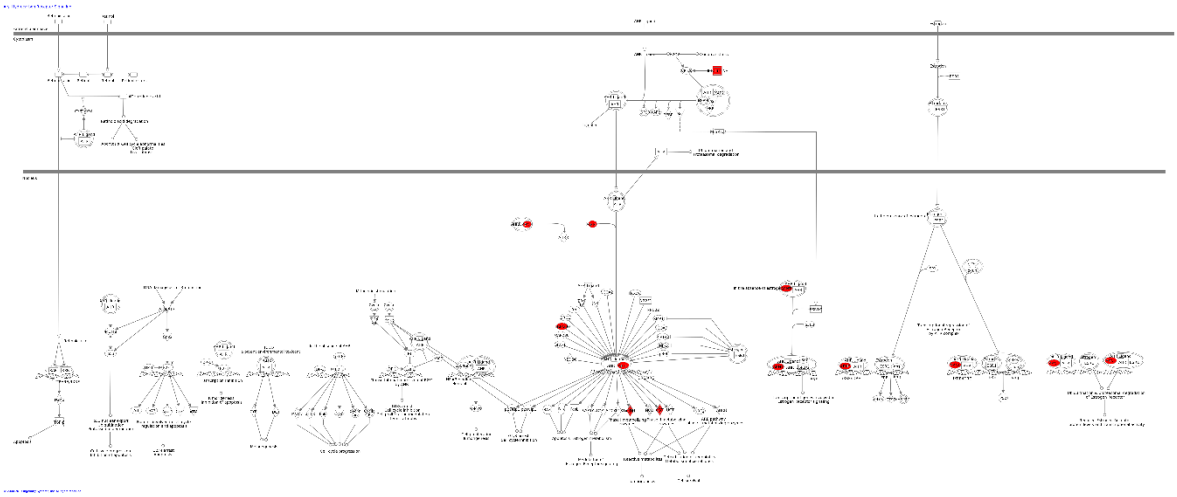

Figure 3-S14 : Causality genes in the Aryl Hydrocarbon Receptor Signaling pathway. Causality genes identified in this study are in red.

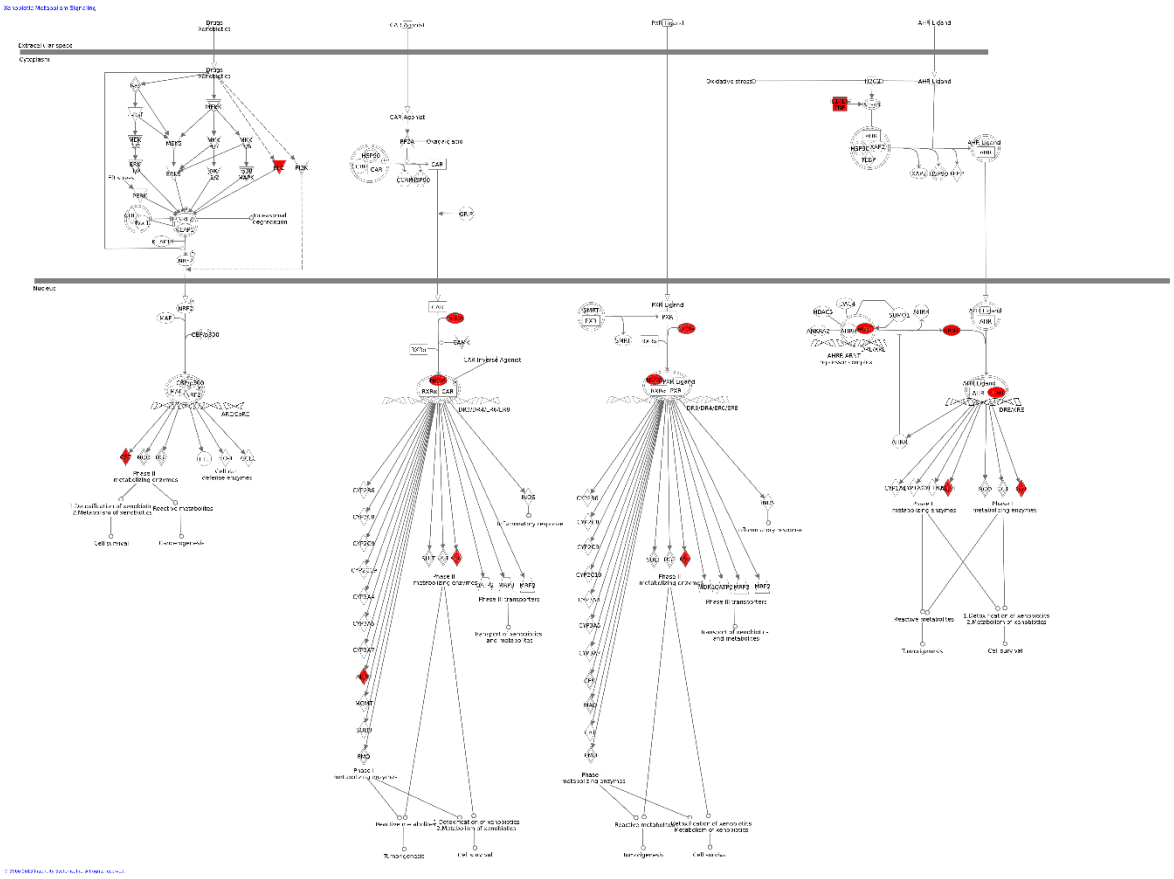

Figure 15 : Causality genes in the Xenobiotic Metabolism Signaling pathway. Causality genes identified in this study are in red.

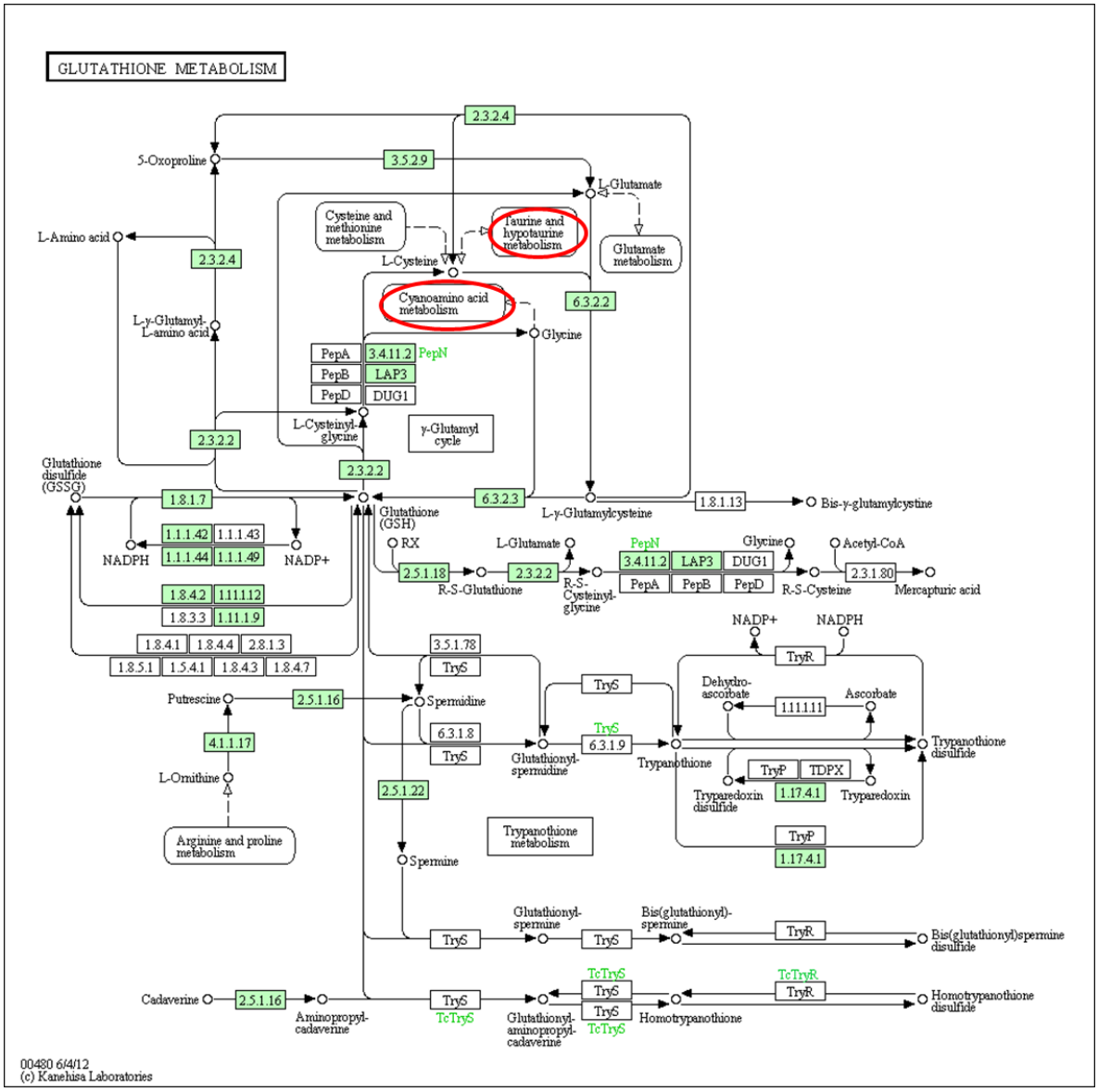

Figure 3-16 : The glutathione metabolism pathway connecting the cyanoamino acid and the taurine/hypotaurine metabolism pathways.

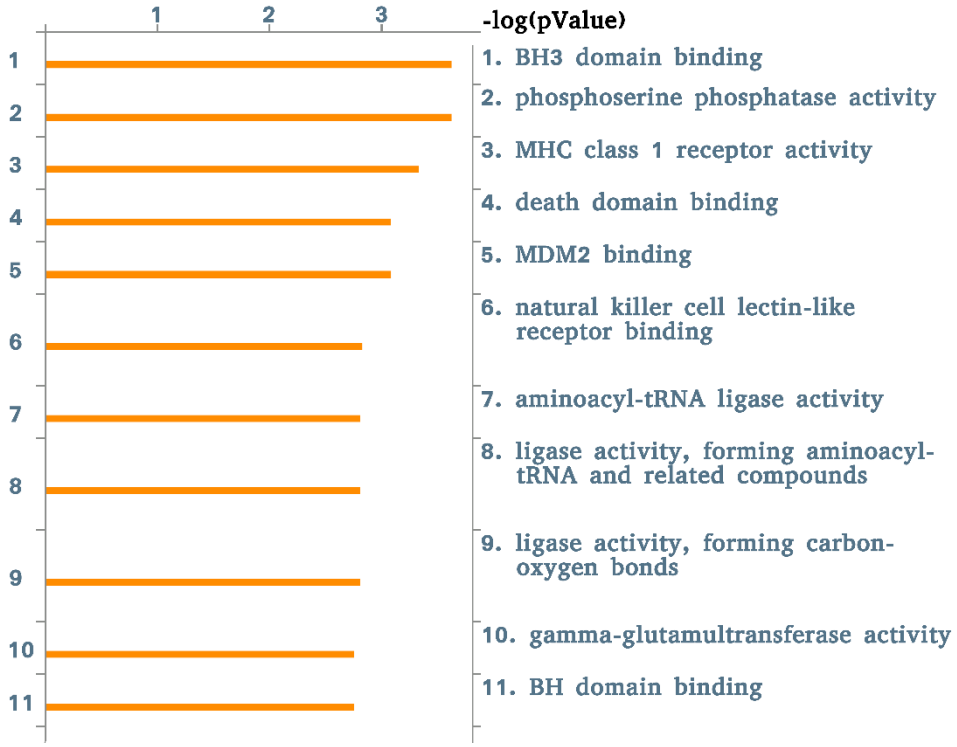

Figure 3-S17 : Gene Ontology (GO) project functional categories enriched for causality genes.

References

1. Hao K, Bosse Y, Nickle DC, et al. Lung eQTLs to Help Reveal the Molecular Underpinnings of Asthma. *PLoS Genet* 2012;8:e1003029.

2. Schadt EE, Lamb J, Yang X, et al. An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet* 2005;37:710-7.

3. Steiling K, van den Berge M, Hijazi K, et al. A dynamic bronchial airway gene expression signature of chronic obstructive pulmonary disease and lung function impairment. *Am J Respir Crit Care Med* 2013;187:933-42.

4. Campbell JD, McDonough JE, Zeskind JE, et al. A gene expression signature of emphysema-related lung destruction and its reversal by the tripeptide GHK. *Genome medicine* 2012;4:67.

5. Soler Artigas M, Loth DW, Wain LV, et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* 2011;43:1082-90.

6. Abboud RT, Vimalanathan S. Pathogenesis of COPD. Part I. The role of protease-antiprotease imbalance in emphysema. *Int J Tuberc Lung Dis* 2008;12:361-7.

7. Butler MW, Fukui T, Salit J, et al. Modulation of cystatin A expression in human airway epithelium related to genotype, smoking, COPD, and lung cancer. *Cancer Res* 2011;71:2572-81.

8. Li J, Dai A, Hu R, Zhu L, Tan S. Positive correlation between PPARgamma/PGC-1alpha and gamma-GCS in lungs of rats and patients with chronic obstructive pulmonary disease. *Acta Biochim Biophys Sin (Shanghai)* 2010;42:603-14.

9. Cho MH, Klanderman BJ, Litonjua AA, Sparrow D, Silverman EK, Raby BA. Folliculin mutations are not associated with severe COPD. *BMC Med Genet* 2008;9:120.

10. Vassallo R, Walters PR, Lamont J, Kottom TJ, Yi ES, Limper AH. Cigarette smoke promotes dendritic cell accumulation in COPD; a Lung Tissue Research Consortium study. *Respir Res* 2010;11:45.

11. Solarewicz-Madejek K, Basinski TM, Crameri R, et al. T cells and eosinophils in bronchial smooth muscle cell death in asthma. *Clin Exp Allergy* 2009;39:845-55.

12. Morissette MC, Vachon-Beaudoin G, Parent J, Chakir J, Milot J. Increased p53 level, Bax/Bcl-x(L) ratio, and TRAIL receptor expression in human emphysema. *Am J Respir Crit Care Med* 2008;178:240-7.

13. Lesseur C, Gilbert-Diamond D, Andrew AS, et al. A case-control study of polymorphisms in xenobiotic and arsenic metabolism genes and arsenic-related bladder cancer in New Hampshire. *Toxicol Lett* 2012;210:100-6.

14. Wilk JB, Walter RE, Laramie JM, Gottlieb DJ, O'Connor GT. Framingham Heart Study genome-wide association: results for pulmonary function measures. *BMC Med Genet* 2007;8 Suppl 1:S8.

15. Yanbaeva DG, Wouters EF, Dentener MA, Spruit MA, Reynaert NL. Association of glutathione-S-transferase omega haplotypes with susceptibility to chronic obstructive pulmonary disease. *Free Radic Res* 2009;43:738-43.

16. Evans DS, Kapahi P, Hsueh WC, Kockel L. TOR signaling never gets old: aging, longevity and TORC1 activity. *Ageing Res Rev* 2011;10:225-37.

17. Kramer EL, Hardie WD, Mushaben EM, et al. Rapamycin decreases airway remodeling and hyperreactivity in a transgenic model of noninflammatory lung disease. *J Appl Physiol* 2011;111:1760-7.

18. Hijikata M, Matsushita I, Tanaka G, et al. Molecular cloning of two novel mucin-like genes in the disease-susceptibility locus for diffuse panbronchiolitis. *Hum Genet* 2011;129:117-28.

19. Kabesch M, Carr D, Weiland SK, von Mutius E. Association between polymorphisms in serine protease inhibitor, kazal type 5 and asthma phenotypes in a large German population sample. *Clin Exp Allergy* 2004;34:340-5.

20. Birben E, Sackesen C, Turgutoglu N, Kalayci O. The role of SPINK5 in asthma related physiological events in the airway epithelium. *Respir Med* 2012;106:349-55.

CHAPITRE 4

Susceptibility genes for lung diseases in the major histocompatibility complex revealed by lung expression quantitative trait loci analysis

Lamontagne M, Joubert P, Timens W, Postma DS, Hao K, Nickle D, Sin DD, Paré PD, Laviolette M, Bossé Y. Eur Respir J. 2016 Aug;48(2):573-6.

Susceptibility genes for lung diseases in the major histocompatibility complex revealed by lung expression quantitative trait loci analysis

Maxime Lamontagne[1], Philippe Joubert[1], Wim Timens[2], Dirkje S. Postma[3], Ke Hao[4], David Nickle[5], Don D. Sin[6,7], Peter D. Pare[6,7], Michel Laviolette[1], Yohan Bossé[1,8]

1) Institut universitaire de cardiologie et de pneumologie de Québec, Québec, Canada

2) University of Groningen, University Medical Center Groningen, Department of Pathology and Medical Biology, GRIAC research institute, Groningen, The Netherlands

3) University of Groningen, University Medical Center Groningen, Department of Pulmonology, GRIAC Research Institute, Groningen, The Netherlands

4) Department of Genetics and Genomics Sciences, Mount Sinai School of Medicine, New York, New York, USA

5) Merck & Co Inc, Rahway, New Jersey, USA

6) University of British Columbia Center for Heart Lung Innovation and Institute for Heart and Lung Health, St Paul's Hospital, Vancouver, British Columbia, Canada

7) Respiratory Division, Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada

8) Department of Molecular Medicine, Laval University, Québec, Canada

Address correspondence to:

Yohan Bossé, Ph.D.

Associate Professor, Laval University

Institut universitaire de cardiologie et de pneumologie de Québec

Pavillon Marguerite-d'Youville, Y2106

2725, chemin Sainte-Foy

Québec (Québec)

Canada, G1V 4G5

Tel: 418-656-8711 ext. 3725

Fax: 418-656-4602

email: yohan.bosse@criucpq.ulaval.ca

Take Home

New susceptibility genes for asthma, idiopathic interstitial pneumonia, lung cancer and lung function in the MHC.

Résumé

Introduction: Le complexe majeur d'histocompatibilité (MHC) a été associé aux maladies respiratoires (asthme, cancer du poumon, fibrose kystique, pneumopathie interstitielle idiopathique) et aux fonctions pulmonaires par des études pangénomiques (GWAS). Par contre, ces études n'ont pu déterminer le ou les gènes causaux.

Objectif : Cette étude vise dans un premier temps à identifier des eQTL pulmonaires dans la région du MHC puis à déterminer le ou les gènes régulés par les variations génétiques identifiées dans les GWAS sur les maladies respiratoires.

Méthodes: Les tissus pulmonaires non-tumoraux de 1,038 patients provenant de trois cohortes ont été analysés dans cette étude. Les eQTL ont d'abord été calculés dans une première cohorte (Laval, n=409) puis répliqués dans deux cohortes supplémentaires (UBC, n=287; Groningen, n=342). Les eQTL pulmonaires ont par la suite été superposés aux résultats des GWAS pour identifier les gènes causaux.

Résultats: 5,790 eQTL pulmonaires ont été identifiés dans le MHC. Parmi les 28 variations génétiques des GWAS précédents sur les maladies pulmonaires et les fonctions respiratoires présentes dans cette région, 14 étaient associées avec l'expression d'un ARNm. Plusieurs gènes identifiés sont biologiquement pertinents et pourraient expliquer les effets des variations génétiques associées à l'asthme (*AGPAT1* et *CDSN*), au cancer du poumon (*BTN3A2* et *ZFP57*), aux fonctions respiratoires (*MICB*) et à la pneumopathie interstitielle idiopathique (*AGPAT1*) dans la région du MHC.

Conclusion: En combinant des eQTL pulmonaires avec les résultats des GWAS précédents, nous avons identifié de possibles gènes causaux pour l'asthme, le cancer du poumon, les fonctions respiratoires et la pneumopathie interstitielle idiopathique dans la région du MHC. Ces résultats nous permettent de mieux comprendre et interpréter les associations génétiques observées pour les maladies respiratoires dans la région du MHC.

Abstract

Background: The Major Histocompatibility Complex (xMHC) has been associated with lung diseases (asthma, cystic fibrosis, idiopathic interstitial pneumonia, lung cancer) and lung function in previous genome-wide association studies (GWAS). As for many loci implicated by GWAS, the causal gene(s) underpinning the genetic association with lung diseases in the xMHC locus remain unknown.

Objective: The goal of this study is first to identify lung eQTLs located inside the major histocompatibility complex and subsequently to reveal the most likely susceptibility genes in this region that mediate the heritable risk of lung diseases identified from previous GWAS.

Methods: Non-tumor lung specimens and DNA from 1,038 patients recruited at three academic sites were included in this study. Lung eQTLs were identified in the discovery set (Laval, n=409) and verified in two replication sets (UBC, n=287; Groningen, n=342). GWAS-SNPs located in the xMHC and associated with lung diseases or lung function were investigated.

Results: A total of 5,790 significant lung eQTLs were detected in the xMHC. Among the 28 GWAS SNPs associated with lung diseases and lung function in the xMHC, 14 were associated with the expression levels of at least one transcript. Potential causal genes that may explain the effect of risk-associated genetic variants were found for asthma (*AGPAT1* and *CDSN*), lung cancer (*BTN3A2* and *ZFP57*), lung function (*MICB*), and idiopathic interstitial pneumonia (*AGPAT1*).

Conclusion: By combining lung eQTLs with results from previous GWAS for lung diseases and traits, this study revealed new susceptibility genes for asthma, idiopathic interstitial pneumonia, lung cancer and lung function within the xMHC. Results of this study represent an important step to elucidate the causal genes underlying the genetic association with lung diseases and lung function in the xMHC region. This study highlights possible causal genes underlying the genetic association with lung diseases and pulmonary function in the xMHC region.

Introduction

The major histocompatibility complex (MHC) has been linked with hundreds of diseases [1]. Genes located in this region play important roles in antigen presentation and the immune response with well-established clinical relevance to blood transfusion, organ transplantation, immunologic and infectious diseases. The MHC is one of the most complex regions of the human genome because of the high gene density, extended linkage disequilibrium (LD) and sequence diversity [2]. An extended region of the MHC (xMHC) of 7.6 Mb containing 400 genes and pseudogenes has been defined [1].

Recent genome-wide association studies (GWAS) have identified that polymorphisms located in the xMHC are associated with lung diseases and related-traits: asthma, cystic fibrosis, idiopathic interstitial pneumonia, lung cancer and lung function (**Table 4-S1**). These diseases/traits are characterized by mechanisms linked to the MHC including persistent inflammation, aberrant immune response and antigen presentation. However, due to the complexity of this locus, limitations of GWAS [3] and tissue-specific characteristic of gene expression [4, 5], the causal genes and genetic mechanisms mediating the heritable risk within this locus remain to be found.

As part of the lung eQTL consortium, we recently reported the discovery of a large number of lung specific eQTLs [6] and identified SNP-mRNA-phenotype triplets associated with airflow obstruction [7]. The present study has two goals. First, to identify lung expression Quantitative Trait Loci (eQTLs) within the xMHC region. Second, identify new susceptibility genes for lung diseases/traits by overlaying lung eQTL results and the xMHC SNPs previously reported to be associated with lung function and respiratory diseases. Susceptibility alleles for respiratory diseases and traits that function as strong lung eQTLs should facilitate the biological interpretation of GWAS results and the identification of causal genes in loci with high gene density and high LD such as the xMHC.

Methods

Subjects

Study subjects and lung specimens have been described previously [6, 8]. Briefly subjects were from three academic sites: Laval University (Quebec, Canada), University of British-Columbia (Vancouver, Canada) and Groningen University (Groningen, The Netherlands), henceforth referred to as Laval, UBC, and Groningen, respectively. At Laval, the lung specimens were provided by the *Institut universitaire de cardiologie et de pneumologie de Québec* (IUCPQ) site of the Respiratory Health Network Biobank of the *Fonds de recherche du Quebec – Santé* (FRQS) (www.tissuebank.ca); at UBC, the lung specimens were provided by the James Hogg Research Center Biobank at St Paul's Hospital, and at Groningen, the lung specimens were provided by the local tissue bank of the Department of Pathology. COPD diagnosis and severity were determined according to the GOLD recommendations [9]. Clinical characteristics of subjects by site are shown in **Table 4-1**. All lung tissue samples were obtained in accordance with Institutional Review Board guidelines at the three sites. Patients provided written informed consent and the study was approved by the ethics committees of the IUCPQ and the UBC-Providence Health Care Research Institute Ethics Board for Laval and UBC, respectively. The study protocol was consistent with the Research Code of the University Medical Center Groningen and Dutch national ethical and professional guidelines (''Code of conduct; Dutch federation of biomedical scientific societies''; http://www.federa.org).

Assays

Genome-wide gene expression and genotyping profiles were obtained using a custom Affymetrix array (GPL10379) and the Illumina Human1M-Duo BeadChip array, respectively. Gene expression data are available through the Gene Expression Omnibus (GEO) repository with the accession number GSE23546. Probe sequences were mapped to the human genome (hg19) using Bowtie [10] and probes not mapping to a coding region or having a common SNP (MAF ≥ 5%) in their sequence were removed. Standard quality controls were performed on genotyping data (MAF ≥ 1%; call rate ≥ 97%; Hardy-Weinberg *P*-value ≥ 0.001). Genotypes were imputed after quality controls with SHAPEIT2 [11] and

IMPUTE v2 [12] using the 1000 Genomes Project Phase 1 data as a reference set. Only subjects that passed genotyping and gene expression quality controls were included in this study. Subjects with missing value for smoking status were also excluded leaving 409, 287 and 342 subjects from Laval, UBC, and Groningen, respectively.

Lung eQTLs by genomic regions

To identify the number of eQTLs by genomic regions, the genome-wide eQTLs results were obtained from our previous report [6]. In that report, we assumed that each transcript could have at most one local-eQTL. Accordingly, all association signals from SNPs within 1 Mb up and downstream of the transcription probe set were considered as a single local-eQTL. The 14,734 local-eQTL-SNPs were aligned with SNPnexus [13] (hg19). The human genome was split in sliding-window size of 20 Mb starting on chromosome 1 and ending at chromosome 22. Windows located at the end of each chromosome were between 10 and 30 Mb. A total of 145 windows were created and each local-eQTL-SNP was assigned to one window.

xMHC locus

The borders of the extended Major Histocompatibility Complex were defined in 2004 by Horton *et al* [1]. The complex is delimited by two genes on chromosome 6: *HIST1H2AA* and *KIFC1*. The expressions of all probe sets located inside this region were used in our analysis, which includes 271 probe sets covering 212 transcripts. 6,872 genotyped SNPs were available in the three datasets.

Statistical analysis

Lung eQTLs in the xMHC region were identified using a different model than our previous genome-wide lung eQTL mapping study [6]. Expression data were adjusted for age, sex, and smoking status using residuals obtained with the robust fitting of linear models function (rlm) in the R statistical package MASS. Residual values deviating from the median by more than three standard deviations were filtered as outliers. Association tests between adjusted expression traits and SNPs were performed using quantitative association tests implemented in PLINK [14] (version 1.90). Association tests were

performed with the "assoc" command and the Wald test asymptotic *P*-values was used. Each possible combination of genotyped SNP and probe set located in the xMHC was tested in the Laval, UBC and Groningen datasets separately. eQTLs were considered significant if they passed Bonferroni correction in Laval (MHC analysis: 0.05 / (271 probe sets x 6872 SNPs), *P*-values $\leq 2.68 \times 10^{-08}$; GWAS Catalog analysis: 0.05 / (271 probe sets x 27 SNPs), *P*-values $\leq 6.83 \times 10^{-06}$) and replicated (*P*-value ≤ 0.05) in UBC and Groningen. Local- and distant-acting eQTLs were defined using a 1 Mb threshold between the SNP and transcriptional start site (TSS) positions of MHC genes. Circular representation of eQTL results were made with Circos [15].

GWAS Catalog

The NHGRI-EBI Catalog of published GWAS was downloaded (https://www.genome.gov/26525384, Accessed 19 March 2015) and hits located in the xMHC were isolated. Five lung diseases/traits were considered: asthma, cystic fibrosis, idiopathic interstitial pneumonia, lung cancer, and lung function. Twenty-nine SNPs were associated with one or more of these diseases/traits with GWAS *P*-values $< 1 \times 10^{-5}$ (**Table 4-S1**). One failed Hardy-Weinberg equilibrium (rs2647044) and six were not available in our genotyped or imputed data. SNPs in perfect LD were found for five of the six missing SNPs using the 1000 Genomes Project browser (http://browser.1000genomes.org/index.html) (**Table 4-S1**). No SNP in perfect LD was found for rs3094548 and the best LD score was with rs238869 (r^2=0.79). Rs2395185, a SNP associated with lung cancer, was serving as a proxy for the cystic fibrosis SNP rs9268905 (r^2=1). Overall, a total of 27 risk-associated genetic variants were considered. Most of the GWAS-SNPs were not in strong LD with each other (**Figure 4-S1**).

Results

Lung eQTLs by genomic regions

Genome-wide eQTL analysis from our previous report revealed 14,734 local-eQTL-SNPs. The number of local-eQTL-SNPs were counted by using a sliding-window size of 20 Mb from chromosome 1 to chromosome 22, except the last window of each chromosome that ranged between 10.3 and 29.3 Mb. **Figure 4-1** and **Table 4-S2** show the number of local-eQTL-SNPs for the 145 windows. The average number of local-eQTL-SNPs among the 145 windows was 101.6 ± 55.1. The three windows with the largest number of local-eQTL-SNPs were chr19:40,000,001–59,128,983 (n=385), chr6:20,000,001-40,000,000 (n=282) and chr19:1-20,000,000 (n=268). The xMHC is included in the interval 20-40 Mb on chromosome 6 and is thus in the window with the second largest number of local-eQTL-SNPs (**Figure 4-1**).

Lung eQTLs in the xMHC

A total of 5,790 significant lung eQTLs were detected in the xMHC (**Table 4-S3**). Please note that in the current xMHC analysis, no restriction was set for the number of local-eQTLs by transcript, which contrasts our previous genome-wide analysis [6] limited to at most one local-eQTL by transcript. The eQTLs consisted of 2,807 SNPs and 66 associated probe sets interrogating 50 transcripts. The majority of eQTLs were local (n= 5,516), but 274 distant-acting eQTLs were found. Local-eQTL-SNPs located closer to the transcriptional start site had lower *P*-values (i.e. higher significance), and explained a greater proportion of the gene expression variance (**Figure 4-S2**). The most significant eQTL detected in the discovery set (Laval) was between *ZFP57* and four SNPs in perfect LD (rs2535238, rs2747421, rs375984, rs387642; *P*-value = 1.06×10^{-124}) (**Figure 4-S3**). eQTL-SNPs were pruned to 663 independent SNPs (LD with $r^2 < 0.8$). **Figure 4-S4** shows results for this subset of local- and distant-eQTLs.

Integration of lung eQTLs with risk-associated SNPs from GWAS on lung diseases

Among the 28 GWAS SNPs associated with lung diseases in the xMHC, 14 were associated with the expression levels of at least one transcript (**Figure 4-2A**; **Table 4-2**; **Table 4-S4**). For all significant results present in **Table 4-2**, the direction of effect was the same in the two replication sets. Nineteen unique eQTLs were found with SNPs previously associated with asthma (6 SNPs; 8 transcripts), two with idiopathic interstitial pneumonia (1 SNP; 2 transcripts), four with lung cancer (2 SNPs; 4 transcripts), 13 with lung function (5 SNPs; 10 transcripts), and none with cystic fibrosis. **Figure 4-2** presents the most relevant lung eQTLs identified for SNPs previously associated with asthma, lung cancer, lung function, and idiopathic interstitial pneumonia. For lung cancer, rs7767099 was associated with mRNA expression of *ZFP57* encoding the transcriptional repressor zinc finger protein 57 (high expression for the risk allele (T)) (**Figure 4-2B**). For asthma, the risk allele (rs9273349-C) was associated with lowered mRNA expression levels of *CDSN* (corneodesmosin) (**Figure 4-2C**). The same risk variant was also associated with lowered expression levels of *AGPAT1* encoding 1-acylglycerol-3-phosphate O-acyltransferase 1 (**Figure 4-2D**). The variant associated with lower FEV_1 and FEV_1/FVC (rs2855812-T) was associated with higher mRNA expression of *MICB* encoding the MHC class I polypeptide-related sequence B (**Figure 4-2E**). Finally, the idiopathic interstitial pneumonia risk allele (rs3132946-A) was associated with higher expression of *AGPAT1* (**Figure 4-2F**). Accordingly, the risk variants for asthma (rs9273349) and idiopathic interstitial pneumonia (rs3132946) have an opposite direction of effect on the expression of *AGPAT1* in lung tissue. The most significant results for cystic fibrosis in the discovery set was between rs2395185 and *CDSN* (*P*-value = 0.0016), which did not reach significance. Three genes were associated with GWAS SNPs for two different traits/diseases: *AGPAT1* and *HLA-DQB1* – asthma and idiopathic interstitial pneumonia, and *ZFP57* – lung cancer and lung function (**Table 4-2**).

Discussion

The goal of this study was first to identify lung eQTLs located inside the major histocompatibility complex and subsequently to reveal the most likely susceptibility genes in this region that mediate the heritable risk of lung diseases and lung function identified from previous GWAS. Whole-genome genotyping and gene expression values of 1,038 lung samples from three independent populations were used to identify lung eQTLs. Analyses were limited to the extended version of the major histocompatibility complex (xMHC). Briefly, 5,790 significant lung eQTLs were discovered within the xMHC and potential causal genes that may explain the effect of risk-associated genetic variants were found for asthma (*AGPAT1* and *CDSN*), lung cancer (*BTN3A2* and *ZFP57*), lung function (*MICB*), and idiopathic interstitial pneumonia (*AGPAT1*).

At the genome-wide scale, the xMHC region had the second highest number of local lung eQTL-SNPs (**Figure 4-1**). The first and third richest eQTL genomic windows were both located on chromosome 19. The identification of the xMHC as one of the richest eQTL regions is not surprising since previous studies have identified many local and distant eQTLs at this locus [16]. Chromosome 19 has the highest gene density of all human chromosomes [17], which may explain the high number of lung eQTLs observed in this study.

Among the 5,790 significant lung eQTLs detected in the xMHC, *ZFP57* was the strongest eQTL-gene in the discovery set (**Table 4-S3**). Four SNPs in perfect LD (rs2535238, rs2747421, rs375984, rs387642; *P*-value = 1.06×10^{-124}) explained 75.5% of the gene expression variance. Interestingly, the association between this gene and rs375984 was made previously in a study on peripheral blood mononuclear cells of 288 healthy subjects [18]. This association was then replicated in purified monocytes from the same subjects. In the Genotype-Tissue Expression project data [4] (http://www.gtexportal.org/home/, Accessed 2 July 2015), the same eQTL is present in 12 tissues including the lung. In this study, the surrogate risk-allele of rs7767099 in LD with the lung cancer-related SNP rs4324798 [19, 20] was associated with higher mRNA expression levels of *ZFP57* (**Figure 4-2B**). *ZFP57* induces *IGF2* expression [21] and high protein expression of *IGF2* is associated with lower survival rate in lung cancer [22]. It is

thus tempting to speculate that rs4324798 mediates the risk of lung cancer by increasing expression levels of *ZFP57* in lung tissue.

One more eQTL gene, *BTN3A2* (butyrophilin, subfamily 3, member A2), was associated with lung cancer-associated SNPs. *BTN3A2* was previously associated with ovarian cancer [23]. Higher expression of *BTN3A2* was associated with higher density of infiltrating T cells, particularly $CD4^+$ cells, and better prognosis. Interestingly, a study by Al-Shibli *et al*. [24] found a positive relation between the number of stromal $CD4^+$ cells and the prognosis of non-small cell lung cancer patients. Whether *BTN3A2* is playing a role in recruiting $CD4^+$ cells in lung cancer will require further studies.

A biologically relevant eQTL-gene associated with an asthma SNP is *CDSN*. This gene has been associated with skin diseases including psoriasis [25] and peeling skin syndrome [26]. Genetic variants in this gene may alter skin barrier contributing to allergic sensitization and allergic asthma akin to the effect associated with SNPs in the filaggrin gene [27]. The other asthma-eQTL-gene identified is *AGPAT1* encoding for an enzyme (lysophosphatidic acid acyltransferase, alpha) that converts lysophosphatidic acid (LPA) into phosphatidic 121acid (PA). The autotaxin-LPA pathway was recently associated with allergic lung inflammation in humans [28].

Lung function measurement, such as FEV1 and FEV1/FVC ratio, are good indicators of population health [29]. In this study, the allele associated with lower FEV_1 and FEV_1/FVC (T) was associated with higher expression of *MICB* (**Figure 4-2E**). The eQTL SNP rs2855812 was nearly significant in a GWAS on lung function (FEV_1 and FEV_1/FVC) [30]. *MICB* encodes a ligand for the NKG2D type II receptor, which is a key mediator of cigarette smoke-stimulated NK cell hyperresponsiveness [31]. Higher levels of *MICB* may predispose individuals to greater inflammation due to the constant activation of the NKG2D receptor.

Idiopathic interstitial pneumonias (IIPs) are diffuse interstitial lung diseases (ILD) of unknown cause characterized by variable degree of interstitial fibrosis and/or infiltration of the interstitial compartment with chronic inflammatory cells. In a recent GWAS of IIPs, rs3132946 was borderline significant [32]. The risk allele of this SNP is associated with

elevated levels of *AGPAT1* in the current study. *AGPAT1* was also associated with an asthma risk variant in this study. The risk allele in IIPs (rs3132946-A) is associated with higher expression levels of *AGPAT1* (**Figure 4-2F**). In asthma, the risk allele (rs9273349-C) is associated with lowered expression of *AGPAT1* (**Figure 4-2D**). These two SNPs are in low LD in our datasets (max r^2 = 0.17). As mentioned above, *AGPAT1* encodes an enzyme that converts LPA into PA. Elevated levels of LPA and/or ATX have been observed in the BAL [33], lung [34] and exhaled breath condensate [35] of IIPs patients. The risk-associated genetic variants for IIPs in the MHCx may thus be mediated through *AGPAT1*.

This study has limitations. eQTLs obtained in this study are derived from non-tumor lung parenchymal samples. Because whole lung tissues, containing multiple cell types, were used, many eQTLs in less prevalent cellular types will not be detected [16]. Even if a large number of eQTLs are shared across tissues [36], a proportion of them are cell type- and tissue-specific [37]. eQTLs are also known to be influenced by disease states. In this study, the majority of patients had lung cancer, but patients with other lung diseases were also included in the replication sets. Due to these limitations, studies in other disease-relevant tissues or cell types and subjects are needed to confirm or refine these results.

In conclusion, integration of lung specific eQTLs located in the xMHC with results from previous GWAS for lung diseases/traits has revealed new susceptibility genes for asthma, lung cancer, lung function and idiopathic interstitial pneumonia. This study provides significant new information concerning our understanding of the complex regulatory mechanism governing the xMHC region and sheds light on the mechanism underlying GWAS loci for lung diseases identified in this region.

Acknowledgements

The authors would like to thank Christine Racine and Sabrina Biardel at the IUCPQ site of the Respiratory Health Network Biobank of the FRQS for their valuable assistance. They also acknowledge the staff at Calcul Québec for IT support with the high-performance computer clusters. At UBC the authors thank the biobank coordinator Dr Mark Elliott. At the Groningen UMCG site Marnix Jonker is thanked for his support in selecting, handling and sending of lung tissues.

Support

The lung eQTL study at Laval University was supported by the Chaire de pneumologie de la Fondation JD Bégin de l'Université Laval, the Fondation de l'Institut universitaire de cardiologie et de pneumologie de Québec, the Respiratory Health Network of the FRQS, the Canadian Institutes of Health Research (MOP - 123369), and the Cancer Research Society and Read for the Cure. MLam is the recipient of a doctoral studentship from the Fonds de recherche Québec - Santé (FRQS). DDS is a Tier 1 Canada Research Chair for COPD. YB was the recipient of a Junior 2 Research Scholar from the FRQS and now holds a Canada Research Chair in Genomics of Heart and Lung Diseases.

References

1. Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK, Lush MJ, Povey S, Talbot CC, Wright MW, Wain HM, Trowsdale J, Ziegler A, Beck S. Gene map of the extended human MHC. *Nat Rev Genet* 2004; 5: 889–899.

2. de Bakker PIW, Raychaudhuri S. Interrogating the major histocompatibility complex with high-throughput genomics. *Hum Mol Genet* 2012; 21: 29–36.

3. Ward LD, Kellis M. Interpreting noncoding genetic variation in complex traits and human disease. *Nat Biotechnol* 2012; 30: 1095–1106.

4. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* 2015; 348: 648–60.

5. Bossé Y. Genome-wide expression quantitative trait loci analysis in asthma. *Curr. Opin. Allergy Clin Immunol* 2013; 13: 487–494.

6. Hao K, Bossé Y, Nickle DC, Paré PD, Postma DS, Laviolette M, Sandford A, Hackett TL, Daley D, Hogg JC, Elliott WM, Couture C, Lamontagne M, Brandsma C-A, van den Berge M, Koppelman G, Reicin AS, Nicholson DW, Malkov V, Derry JM, Suver C, Tsou J a., Kulkarni A, Zhang C, Vessey R, Opiteck GJ, Curtis SP, Timens W, Sin DD. Lung eQTLs to Help Reveal the Molecular Underpinnings of Asthma. *PLoS Genet* 2012; 8: e1003029.

7. Lamontagne M, Timens W, Hao K, Bossé Y, Laviolette M, Steiling K, Campbell JD, Couture C, Conti M, Sherwood K, Hogg JC, Brandsma C-A, Berge M van den, Sandford A, Lam S, Lenburg ME, Spira A, Paré PD, Nickle D, Sin DD, Postma DS. Genetic regulation of gene expression in the lung identifies CST3 and CD22 as potential causal genes for airflow obstruction. *Thorax* 2014; 69: 997–1004.

8. Bossé Y, Postma DS, Sin DD, Lamontagne M, Couture C, Gaudreault N, Joubert P, Wong V, Elliott M, Van Den Berge M, Brandsma C a., Tribouley C, Malkov V, Tsou J a., Opiteck GJ, Hogg JC, Sandford AJ, Timens W, Paré PD, Laviolette M. Molecular Signature of Smoking in Human Lung Tissues. *Cancer Res* 2012; 72: 3753–3763.

9. Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, Fukuchi Y, Jenkins C, Rodriguez-Roisin R, Van Weel C, Zielinski J. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2007; 176: 532–555.

10. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009; 10: R25.

11. O'Connell J, Gurdasani D, Delaneau O, Pirastu N, Ulivi S, Cocca M, Traglia M, Huang J, Huffman JE, Rudan I, McQuillan R, Fraser RM, Campbell H, Polasek O, Asiki G, Ekoru K, Hayward C, Wright AF, Vitart V, Navarro P, Zagury JF, Wilson JF, Toniolo D, Gasparini P, Soranzo N, Sandhu MS, Marchini J. A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness. *PLoS Genet* 2014; 10: e1004234.

12. Howie BN, Donnelly P, Marchini J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genet* 2009; 5: e1000529.

13. Dayem Ullah AZ, Lemoine NR, Chelala C. SNPnexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update). *Nucleic Acids Res* 2012; July: W65–W70.

14. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015; 4: 7.

15. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra M a. Circos: an information aesthetic for comparative genomics. *Genome Res* 2009; 19: 1639–1645.

16. Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A, Ellis P, Langford C, Vannberg FO, Knight JC. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat Genet* 2012; 44: 502–510.

17. Grimwood J, Gordon L a, Olsen A, Terry A, Schmutz J, Lamerdin J, Hellsten U, Goodstein D, Couronne O, Tran-Gyamfi M, Aerts A, Altherr M, Ashworth L, Bajorek E, Black S, Branscomb E, Caenepeel S, Carrano A, Caoile C, Chan YM, Christensen M, Cleland C a, Copeland A, Dalin E, Dehal P, Denys M, Detter JC, Escobar J, Flowers D, Fotopulos D, et al. The DNA sequence and biology of human chromosome 19. *Nature* 2004; 428: 529–535.

18. Plant K, Fairfax BP, Makino S, Vandiedonck C, Radhakrishnan J, Knight JC. Fine mapping genetic determinants of the highly variably expressed MHC gene ZFP57. *Eur J Hum Genet* 2014; 22: 568–571.

19. Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, Mirabello L, Jacobs K, Wheeler W, Yeager M, Bergen AW, Li Q, Consonni D, Pesatori AC, Wacholder S, Thun M, Diver R, Oken M, Virtamo J, Albanes D, Wang Z, Burdette L, Doheny KF, Pugh EW, Laurie C, Brennan P, Hung R, Gaborieau V, McKay JD, Lathrop M, et al. A Genome-wide Association Study of Lung Cancer Identifies a Region of Chromosome 5p15 Associated with Risk for Adenocarcinoma. *Am J Hum Genet* 2009; 85: 679–691.

20. Wang Y, Broderick P, Webb E, Wu X, Vijayakrishnan J, Matakidou A, Qureshi M, Dong Q, Gu X, Chen WV, Spitz MR, Eisen T, Amos CI, Houlston RS. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat Genet* 2008; 40: 1407–1409.

21. Tada Y, Yamaguchi Y, Kinjo T, Song X, Akagi T, Takamura H, Ohta T, Yokota T, Koide H. The stem cell transcription factor ZFP57 induces IGF2 expression to promote anchorage-independent growth in cancer cells. *Oncogene* 2014; 34: 752–760.

22. Takanami I, Imamuma T, Hashizume T, Kikuchi K, Yamamoto Y, Yamamoto T, Kodaira S. Insulin-like growth factor-II as a prognostic factor in pulmonary adenocarcinoma. *J Surg Oncol* 1996; 61: 205–208.

23. Le Page C, Marineau A, Bonza PK, Rahimi K, Cyr L, Labouba I, Madore J, Delvoye N, Mes-Masson AM, Provencher DM, Cailhier JF. BTN3A2 expression in

epithelial ovarian cancer is associated with higher tumor infiltrating T cells and a better prognosis. *PLoS One* 2012; 7: e38541.

24. Al-Shibli KI, Donnem T, Al-Saad S, Persson M, Bremnes RM, Busund LT. Prognostic effect of epithelial and stromal lymphocyte infiltration in non-small cell lung cancer. *Clin Cancer Res* 2008; 14: 5220–5227.

25. Orrù S, Giuressi E, Carcassi C, Casula M, Contu L. Mapping of the major psoriasis-susceptibility locus (PSORS1) in a 70-Kb interval around the corneodesmosin gene (CDSN). *Am J Hum Genet* 2005; 76: 164–171.

26. Israeli S, Zamir H, Sarig O, Bergman R, Sprecher E. Inflammatory peeling skin syndrome caused by a mutation in CDSN encoding corneodesmosin. *J Invest Dermatol* 2011; 131: 779–781.

27. Hudson TJ. Skin barrier function and allergic risk. *Nat Genet* 2006; 38: 399–400.

28. Park GY, Lee YG, Berdyshev E, Nyenhuis S, Du J, Fu P, Gorshkova I a., Li Y, Chung S, Karpurapu M, Deng J, Ranjan R, Xiao L, Jaffe HA, Corbridge SJ, Kelly E a B, Jarjour NN, Chun J, Prestwich GD, Kaffe E, Ninou I, Aidinis V, Morris AJ, Smyth SS, Ackerman SJ, Natarajan V, Christman JW. Autotaxin production of lysophosphatidic acid mediates allergic asthmatic inflammation. *Am J Respir Crit Care Med* 2013; 188: 928–940.

29. Myint PK, Luben RN, Surtees PG, Wainwright NWJ, Welch a. a., Bingham S a., Wareham NJ, Day NE, Khaw KT. Respiratory function and self-reported functional health: EPIC-Norfolk population study. *Eur Respir J* 2005; 26: 494–502.

30. Soler Artigas M, Loth D, Wain L, Gharib S, Obeidat M, Tang W, Zhai G, Zhao J, Smith A, Huffman J, Albrecht E, Jackson C, Evans D, Cadby G, Fornage M, Manichaikul A, Lopez L, Johnson T, Aldrich M, Aspelund T, Barroso I, Campbell H, Cassano P, Couper D, Eiriksdottir G, Franceschini N, Garcia M, Gieger C, Gislason G, Grkovic I, et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet* 2011; 43: 1082–1090.

31. Wortham BW, Eppert BL, Motz GT, Flury JL, Orozco-Levi M, Hoebe K, Panos RJ, Maxfield M, Glasser SW, Senft AP, Raulet DH, Borchers MT. NKG2D mediates NK cell hyperresponsiveness and influenza-induced pathologies in a mouse model of chronic obstructive pulmonary disease. *J Immunol* 2012; 188: 4468–4475.

32. Fingerlin TE, Murphy E, Zhang W, Peljto a L, Brown KK, Steele MP, Loyd JE, Cosgrove GP, Lynch D, Groshong S, Collard HR, Wolters PJ, Bradford WZ, Kossen K, Seiwert SD, du Bois RM, Garcia CK, Devine MS, Gudmundsson G, Isaksson HJ, Kaminski N, Zhang Y, Gibson KF, Lancaster LH, Cogan JD, Mason WR, Maher TM, Molyneaux PL, Wells a U, Moffatt MF, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet* 2013; 45: 613–620.

33. Tager AM, LaCamera P, Shea BS, Campanella GS, Selman M, Zhao Z, Polosukhin V, Wain J, Karimi-Shah B a, Kim ND, Hart WK, Pardo A, Blackwell TS, Xu Y, Chun J, Luster AD. The lysophosphatidic acid receptor LPA1 links pulmonary fibrosis to lung injury by mediating fibroblast recruitment and vascular leak. *Nat Med* 2008; 14: 45–54.

34. Oikonomou N, Mouratis MA, Tzouvelekis A, Kaffe E, Valavanis C, Vilaras G, Karameris A, Prestwich GD, Bouros D, Aidinis V. Pulmonary autotaxin expression contributes to the pathogenesis of pulmonary fibrosis. *Am J Respir Cell Mol Biol* 2012; 47: 566–574.

35. Montesi SB, Mathai SK, Brenner LN, Gorshkova I a, Berdyshev E V, Tager AM, Shea BS. Docosatetraenoyl LPA is elevated in exhaled breath condensate in idiopathic pulmonary fibrosis. *BMC Pulm Med* 2014; 14: 1–7.

36. Grundberg E, Small KS, Hedman AK, Nica AC, Buil A, Keildson S, Bell JT, Yang T-P, Meduri E, Barrett A, Nisbett J, Sekowska M, Wilk A, Shin S-Y, Glass D, Travers M, Min JL, Ring S, Ho K, Thorleifsson G, Kong A, Thorsteindottir U, Ainali C, Dimas AS, Hassanali N, Ingle C, Knowles D, Krestyaninova M, Lowe CE, Di Meglio P, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet* 2012; 44: 1084–1089.

37. Fu J, Wolfs MGM, Deelen P, Westra H-J, Fehrmann RSN, Te Meerman GJ, Buurman W a, Rensen SSM, Groen HJM, Weersma RK, van den Berg LH, Veldink J, Ophoff R a, Snieder H, van Heel D, Jansen RC, Hofker MH, Wijmenga C, Franke L. Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. *PLoS Genet* 2012; 8: e1002431.

Tables

Table 4-1 : Clinical characteristics of patients that passed gene expression and genotyping quality control filters

	Laval (n = 409)	UBC (n = 287)	Groningen (n = 342)
Male (%)	229 (56.0)	132 (54.0)	182 (53.2)
Age (years)	63.3 ± 9.9	61.6 ± 13.2	51.5 ± 15.6
FEV1 predicted (%)	82.2 ± 17.5 [10]	83.4 ± 22.8 [33]	45.6 ± 30.4 [39]
FVC predicted (%)	91.7 ± 15.1 [25]	91.3 ± 19.0 [31]	65.1 ± 26.4 [57]
FEV1/FVC	0.67 ± 0.10 [26]	0.68 ± 0.12 [33]	0.54 ± 0.22 [34]
COPD	211 (51.6 %) [34]	108 (37.6 %) [56]	155 (45.3 %) [104]
Stage 1 : Mild	82 (38.9 %)	42 (38.9 %)	20 (12.9 %)
Stage 2 : Moderate	117 (55.4 %)	87 (52.8 %)	38 (24.5 %)
Stage 3 : Severe	11 (5.2 %)	2 (1.9 %)	19 (12.3 %)
Stage 4 : Very Severe	1 (0.5 %)	7 (6.4 %)	67 (43.2 %)
NA	0	0	11 (7.1 %)
Smoking			
Smoker	90 (22.0 %)	98 (34.1 %)	57 (16.6 %)
Ex-Smoker	283 (69.2 %)	163 (56.8 %)	185 (54.0 %)
Non-Smoker	36 (8.8 %)	26 (9.1 %)	100 (29.4 %)
Pack-years in ever-smokers	48.5 ± 27.4 [35]	44.7 ± 28.8 [12]	31.2 ± 17.4 [30]

FEV1 : forced expiratory volume in 1 second.

FVC : forced vital capacity.

[-] = missing value.

Table 4-2 : Significant lung eQTLs detected in the xMHC for GWAS associated SNPs with lung diseases

Phenotype	GWAS SNP	Analysed SNP	eQTL-gene	Laval P	UBC P	Groningen P	Effect of risk allele on gene expression
Asthma							
	rs987870	rs987870	AB209058	$1.06x10^{-54}$	$1.61x10^{-35}$	$3.40x10^{-33}$	Down
	rs9273349	rs9273349	*HLA-DQB1*	$3.96x10^{-12}$	$3.20x10^{-11}$	$1.33x10^{-15}$	Down
	rs9273373	rs1063355	*HLA-DQB1*	$3.96x10^{-12}$	$3.20x10^{-11}$	$1.33x10^{-15}$	Down
	rs9272346	rs9272346	*HLA-DQB1*	$1.77x10^{-11}$	$8.95x10^{-11}$	$8.57x10^{-15}$	Down
	rs987870	rs987870	*HLA-DPB1*	$3.71x10^{-11}$	$5.53x10^{-5}$	$9.10x10^{-13}$	Up
	rs9272346	rs9272346	*HLA-DQB2*	$1.94x10^{-10}$	0.037	$4.82x10^{-4}$	Down
	rs9273349	rs9273349	*HLA-DQB2*	$2.73x10_{-10}$	0.027	$3.14x10^{-4}$	Down
	rs9273373	rs1063355	*HLA-DQB2*	$2.73x10_{-10}$	0.027	$3.14x10^{-4}$	Down
	rs9500927	rs9500927	AB209058	$5.85x10_{-10}$	$5.35x10^{-7}$	$2.78x10^{-5}$	Down
	rs9273349	rs9273349	*AGPAT1*	$7.74x10_{-9}$	$1.61x10^{-11}$	$1.53x10^{-11}$	Down
	rs9273373	rs1063355	*AGPAT1*	$7.74x10_{-9}$	$1.61x10^{-11}$	$1.53x10^{-11}$	Down
	rs9272346	rs9272346	*AGPAT1*	$1.27x10_{-8}$	$1.24x10^{-11}$	$1.73x10^{-10}$	Down
	rs9273349	rs9273349	*CDSN*	$4.50x10_{-7}$	0.0089	0.0039	Down
	rs9273373	rs1063355	*CDSN*	$4.50x10_{-7}$	0.0089	0.0039	Down
	rs9273349	rs9273349	*TAP2*	$5.54x10_{-7}$	0.0030	0.041	Down
	rs9273373	rs1063355	*TAP2*	$5.54x10^{-7}$	0.0030	0.041	Down
	rs9272346	rs9272346	*CDSN*	$1.79x10^{-6}$	0.0066	0.0043	Down
	rs987870	rs987870	*HLA-DPB2*	$2.26x10^{-6}$	0.0081	$3.09x10^{-7}$	Up
	rs9275698	rs9275698	*TAP2*	$2.30x10^{-6}$	0.0021	$7.11x10^{-4}$	Up
Idiopathic interstitial pneumonia							
	rs3132946	rs3132946	*HLA-DQB1*	$2.21x10^{-10}$	$4.12x10^{-8}$	$4.58x10^{-8}$	Up
	rs3132946	rs3132946	*AGPAT1*	$1.12x10^{-9}$	$2.79x10^{-8}$	$4.60x10^{-7}$	Up
Lung cancer							
	rs4324798	rs7767099	*BTN3A2*	$1.92x10^{-15}$	$3.60x10^{-24}$	$5.73x10^{-30}$	Down
	rs4324798	rs7767099	*ZFP57*	$3.36x10^{-10}$	$5.03x10^{-17}$	$4.95x10^{-18}$	Up
	rs3117582	rs3117582	D83543	$3.57x10^{-8}$	0.0012	$1.16x10^{-8}$	Up
	rs4324798	rs7767099	BC035828 / BC035659	$3.73x10^{-6}$	$9.44x10^{-13}$	$4.22x10^{-12}$	Up
Lung function							
	rs6903823	rs6903823	*ZSCAN31*	$3.22x10^{-12}$	$2.59x10^{-7}$	0.0037	Up
	rs2252711	rs2256266	*IFITM4P*	$1.44x10^{-11}$	$5.02x10^{-17}$	$7.44x10^{-9}$	Down

	rs3094548	rs238869	*ZFP57*	$1.60x10^{-9}$	$3.85x10^{-10}$	$1.53x10^{-12}$	Down
	rs6903823	rs6903823	*HCG4B*	$4.73x10^{-8}$	$9.31x10^{-6}$	$2.08x10^{-8}$	Up
	rs2252711	rs2256266	AX748153 / AK093478	$1.08x10^{-7}$	$6.12x10^{-6}$	$3.20x10^{-4}$	Down
	rs2252711	rs2256266	*ZFP57*	$5.97x10^{-7}$	$8.93x10^{-8}$	$3.36x10^{-12}$	Up
	rs2252711	rs2256266	*HCG4B*	$6.80x10^{-7}$	$1.12x10^{-8}$	$4.06x10^{-7}$	Up
	rs2252711	rs2256266	*HLA-J*	$8.60x10^{-7}$	0.0044	$5.90x10^{-4}$	Down
	rs2252711	rs2256266	*HLA-F*	$1.05x10^{-6}$	$3.27x10^{-5}$	$3.06x10^{-12}$	Up
	rs2855812	rs2855812	*MICB*	$1.20x10^{-6}$	0.0082	$2.17x10^{-4}$	Up
	rs2857595	rs2857595	*PSORS1C1*	$1.21x10^{-6}$	$1.12x10^{-5}$	$1.99x10^{-5}$	Up
	rs2252711	rs2256266	*HLA-A*	$1.37x10^{-6}$	$1.30x10^{-4}$	$8.06x10^{-6}$	Down
	rs3094548	rs238869	*HLA-A*	$3.80x10^{-6}$	$7.09x10^{-11}$	$1.12x10^{-14}$	Up

For transcripts with many significant probe sets, only the probe set with the smaller P-value in Laval is presented.

Figures

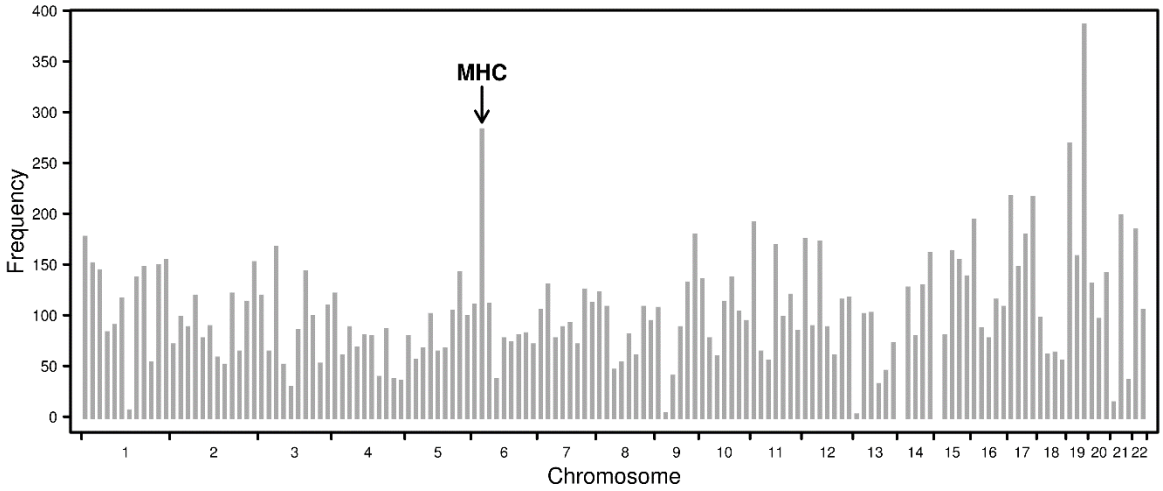

Figure 4-1: Number of local-eQTL-SNPs per sliding-window of 20 Mb.

The x-axis represents windows of 20 Mb starting at chromosome 1 and ending at chromosome 22. The y-axis is the number of local-eQTL-SNPs identified in each window. The xMHC is contained in the window with the second largest number of lung eQTLs.

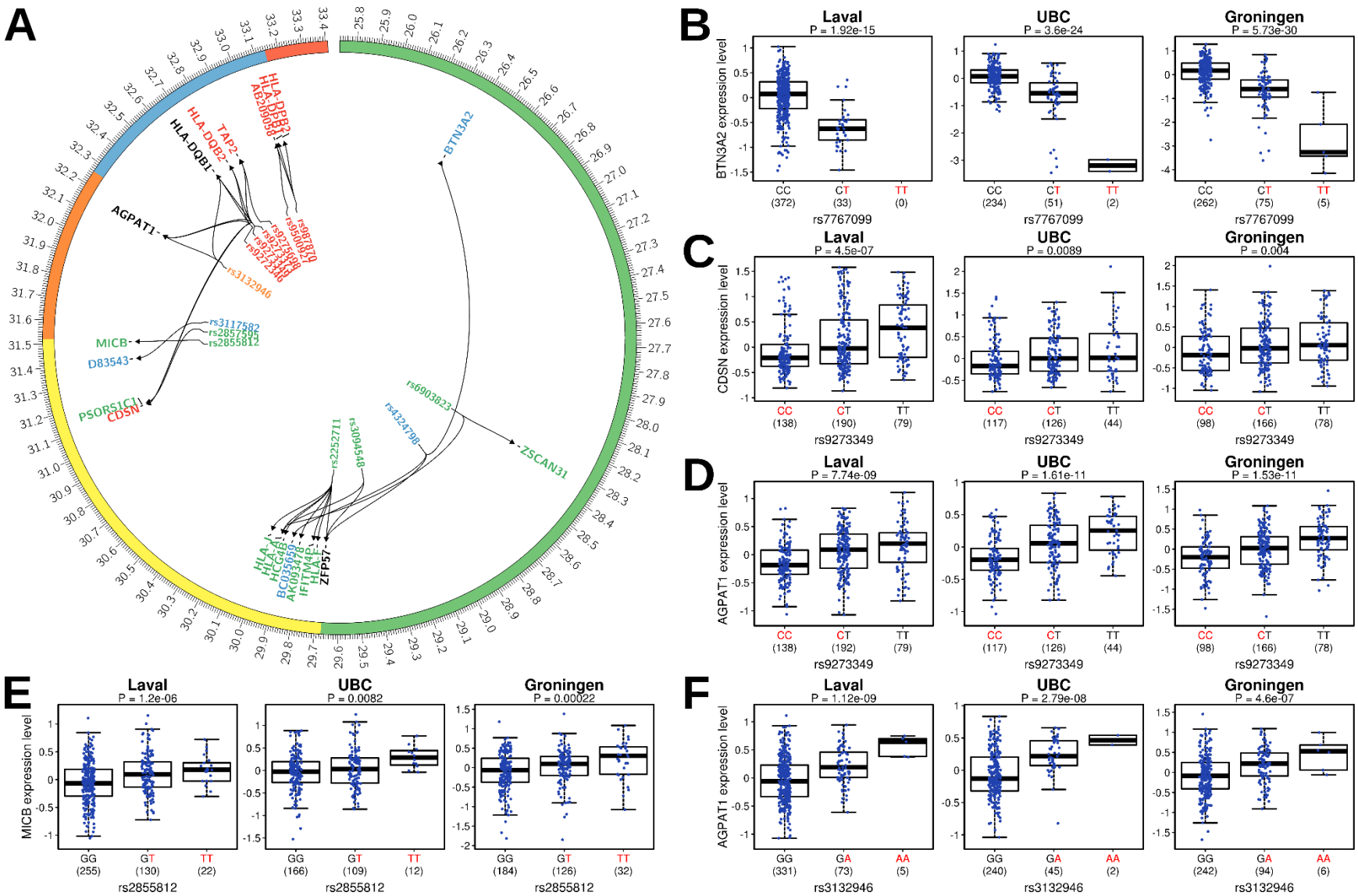
A
B
C
D
E
F
Laval
UBC
Groningen
P = 1.92e-15
P = 3.6e-24
P = 5.73e-30
BTN3A2 expression level
rs7767099
P = 4.5e-07
P = 0.0089
P = 0.004
CDSN expression level
rs9273349
P = 7.74e-09
P = 1.61e-11
P = 1.53e-11
AGPAT1 expression level
P = 1.2e-06
P = 0.0082
P = 0.00022
MICB expression level
rs2855812
P = 1.12e-09
P = 2.79e-08
P = 4.6e-07
rs3132946
BTN3A2
ZSCAN31
AGPAT1
MICB
D83543
PSORS1C1
CDSN
ZFP57
HLA-DQB1
HLA-DQB2
TAP2
rs3132946
rs3117582
rs2857595
rs2855812
rs6903823
rs4324798
rs3094548
rs2252711

Figure 4-2 : Lung eQTL in the xMHC for GWAS-nominated SNPs of lung diseases and traits.

A) Graphical representation of significant lung eQTLs identified in the xMHC considering only GWAS associated SNPs for lung diseases (asthma, lung cancer, interstitial lung disease) and lung function. The outer border of the circle is the chromosomal position in Mb. xMHC sub-regions are color coded (green - extended class I; yellow - classical class I; orange - classical class III; blue - classical class II; red - extended class II). SNPs and transcripts are color coded according to their GWAS associated trait or disease (green – pulmonary function; blue – lung cancer; orange – interstitial lung disease; red – asthma). Transcripts in black are associated with GWAS SNPs of two different traits or diseases. Black arrow heads represent the location of eQTL-genes. B-F) Boxplots of gene expression levels in the lung according to genotype groups. The y-axis shows the mRNA expression levels. The x-axis represents the three genotype groups for the SNP with the number of individuals in parenthesis. The red allele is the risk allele identified in GWAS. Box boundaries, whiskers and centre mark in boxplots represent the first and third quartiles, the most extreme data point which is no more than 1.5 times the IQR, and median, respectively. The left, centre and right panels show the results for Laval, University of British Columbia (UBC) and Groningen samples, respectively. B) Lung cancer-associated SNP rs7767099 – ZFP57. C) Asthma-associated SNP rs9273349 – CDSN. D) Asthma-associated SNP rs9273349 – AGPAT1. E) Lung function-associated SNP rs2855812 – MICB. F) Interstitial lung disease-associated

Supplementary data

Tables

Table 4-S1 : GWAS SNPs with P-values < 1 x 10-5

See the attached file.

Table 4-S2 : Number of local-eQTL-SNPs per window of 20 Mb.

See the attached file.

Figures

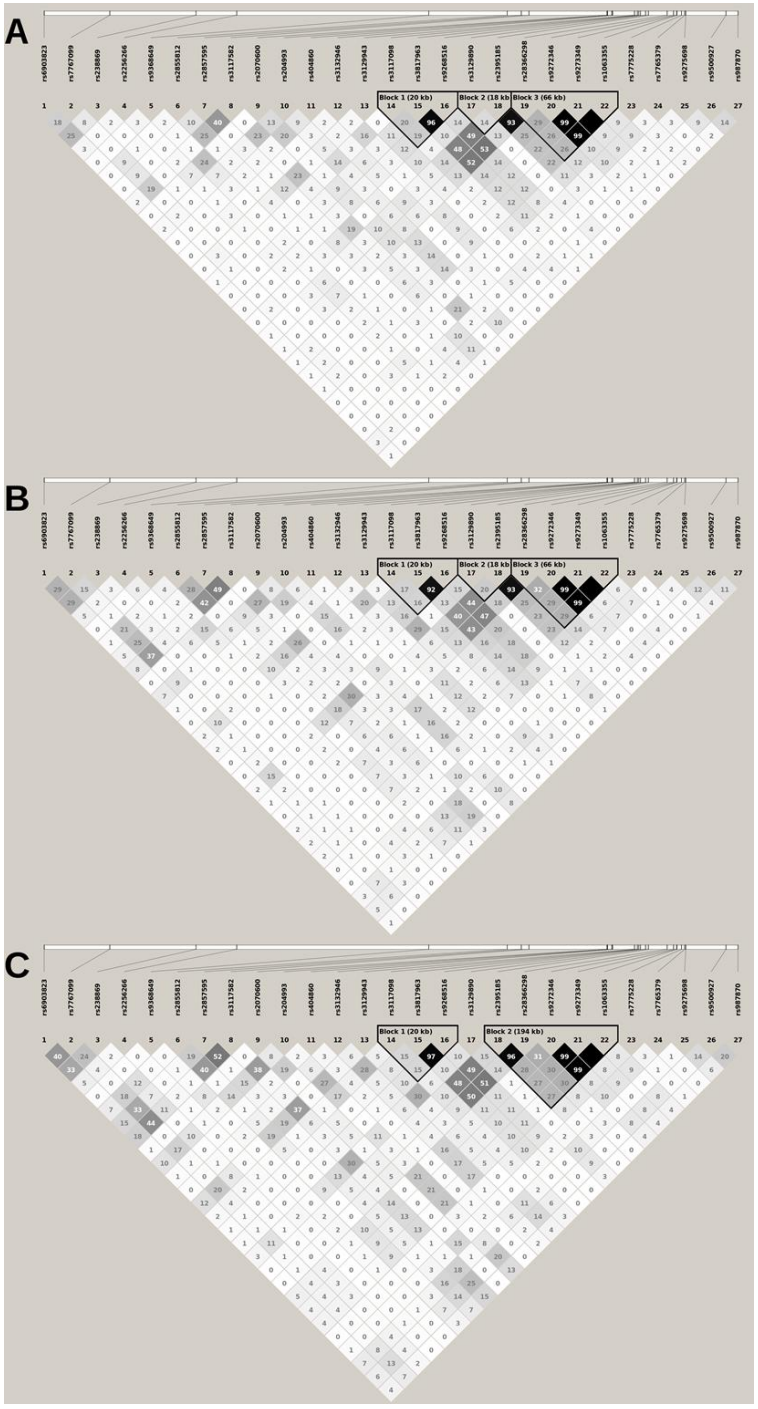

Figure 4-S1 : Linkage disequilibrium plot of selected GWAS SNPs.

A) Laval. B) UBC. C) Groningen.

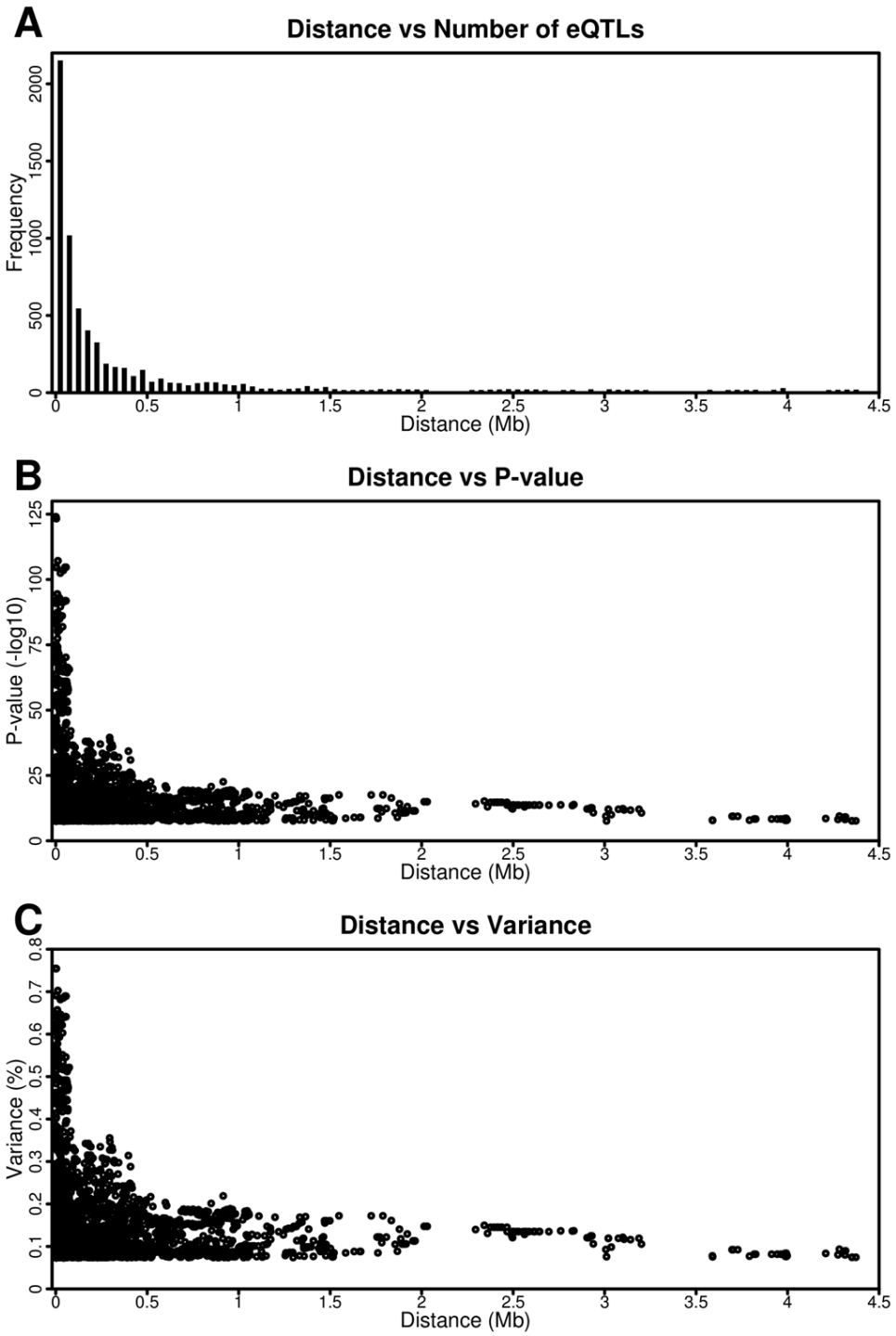

Figure 4-S2 : The effect of distance between the eQTL-SNP and the transcription start site (TSS) of the eQTL-gene in the discovery cohort (Laval).

The x-axis is the distance in Mb between the SNPs and the TSS of the associated gene. A) The number of eQTLs by section of 50 kb. B) eQTL P-values in the discovery set. C) The variance explained by each eQTL.

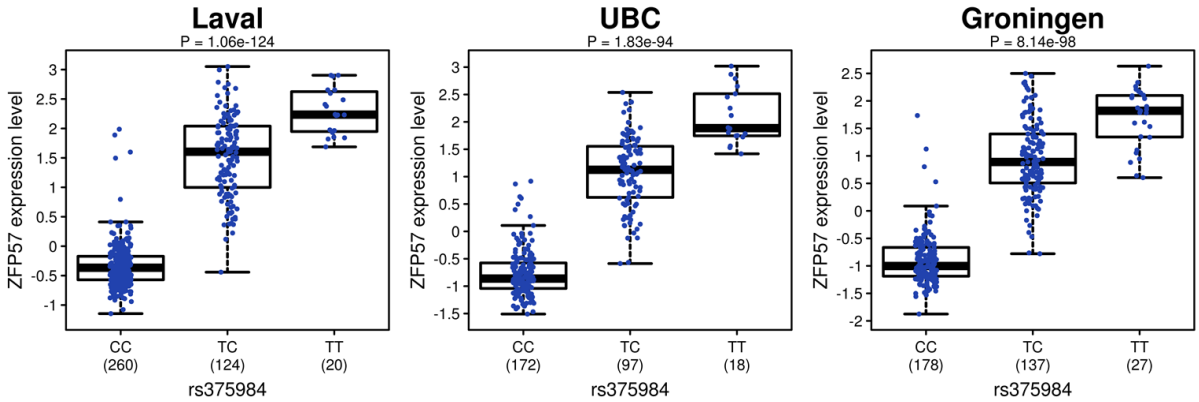

Figure 4-S3 : Boxplots of ZFP57 gene expression levels in the lung according to genotype groups of rs375984.

Boxplots of gene expression levels in the lung according to genotype groups. The y-axis shows the mRNA expression levels of *ZFP57*. The x-axis represents the three genotype groups for rs375984 with the number of individuals in parenthesis. Box boundaries, whiskers and centre mark in boxplots represent the first and third quartiles, the most extreme data point which is no more than 1.5 times the IQR, and median, respectively. The left, centre and right panels show the results for Laval, University of British Columbia (UBC) and Groningen samples, respectively.

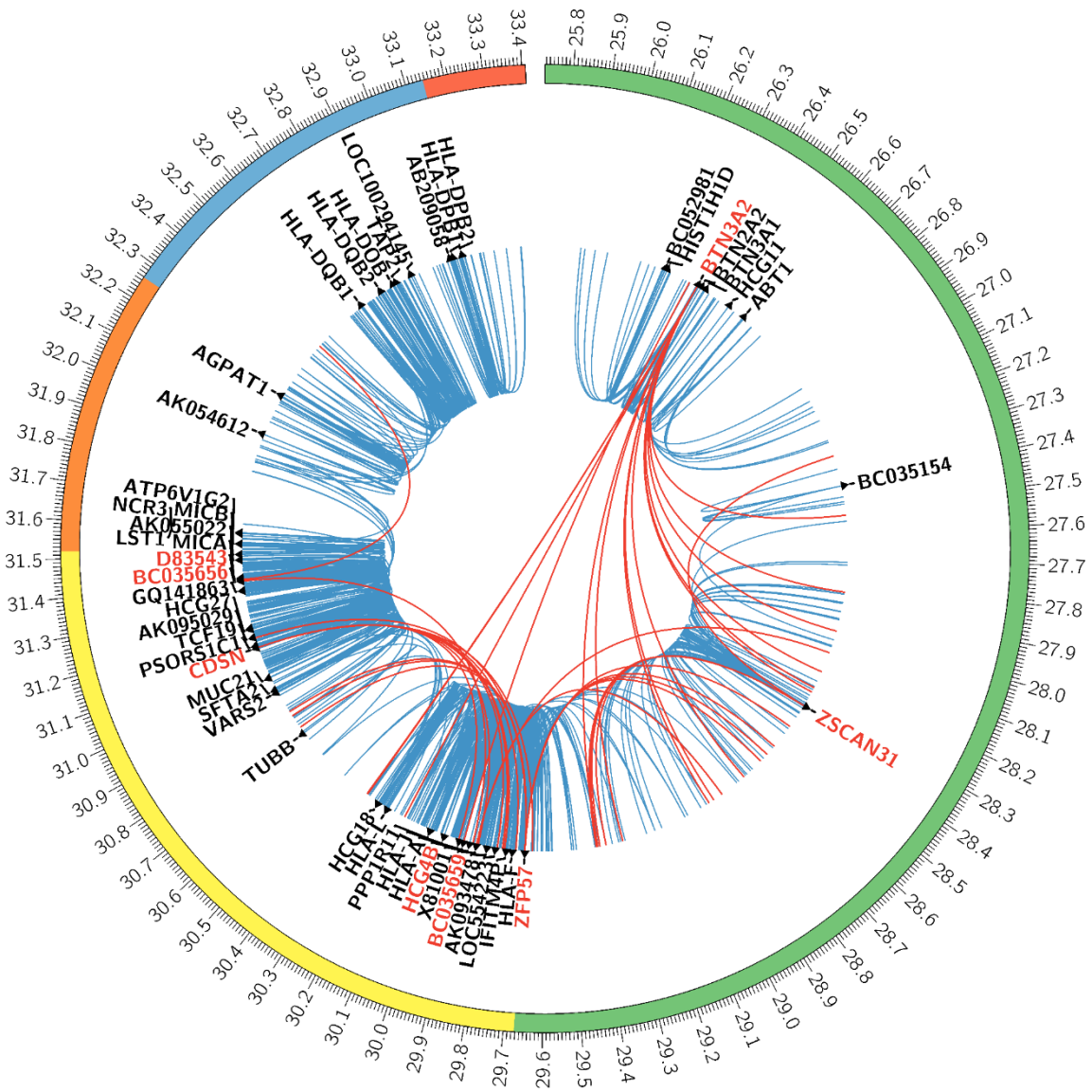

Figure 4-S4 : Circos plot of lung eQTLs located in the xMHC.

Graphical representation of significant lung eQTLs identified in the xMHC considering only independent SNPs (LD with $r^2 < 0.8$). The outer border of the circle is the chromosomal position in Mb. xMHC sub-regions are color coded (green - extended class I; yellow - classical class I; orange - classical class III; blue - classical class II; red - extended class II). All transcripts present (n=50) are associated with at least one local-eQTL-SNP. Transcripts in red (n=8) are also associated with distant-eQTL-SNPs. Blue lines are local-eQTLs (n=1,262) and red lines are distant-eQTLs (n=47). Black arrow heads represent the location of eQTL-gene.

CHAPITRE 5

Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis

Brian D. Hobbs, Kim de Jong, **Maxime Lamontagne**, Yohan Bossé, Nick Shrine, María Soler Artigas, Louise V. Wain, Ian P. Hall, Victoria E. Jackson, Annah B. Wyss, Stephanie J. London, Kari E. North, Nora Franceschini, David P. Strachan, Terri H. Beaty, John E. Hokanson, James D. Crapo, Peter J. Castaldi, Robert P. Chase, Traci M. Bartz, Susan R. Heckbert, Bruce M. Psaty, Sina A. Gharib, Pieter Zanen, Jan W. Lammers, Matthijs Oudkerk, H. J. Groen, Nick Locantore, Ruth Tal-Singer, Stephen I. Rennard, Wim Timens, Peter D. Paré, Jeanne C. Latourelle, Josée Dupuis, George T. O'Connor, Jemma B. Wilk, Woo Jin Kim, Mi Kyeong Lee, Yeon-Mok Oh, Judith M. Vonk, Harry J. de Koning, Shuguang Leng, Steven A. Belinsky, Yohannes Tesfaigzi, Ani Manichaikul, Xin-Qun Wang, Stephen S. Rich, R Graham Barr, David Sparrow, Augusto L. Litonjua, Per Bakke, Amund Gulsvik, Lies Lahousse, Guy G. Brusselle, Bruno H. Stricker, André G. Uitterlinden, Elizabeth J. Ampleford, Eugene R. Bleecker, Prescott G. Woodruff, Deborah A. Meyers, Dandi Qiao, David A. Lomas, Jae-Joon Yim, Deog Kyeom Kim, Iwona Hawrylkiewicz, Pawel Sliwinski, Megan Hardin, Tasha E. Fingerlin, David A. Schwartz, Dirkje S. Postma, William MacNee, Martin D. Tobin, Edwin K. Silverman, H. Marike Boezen, Michael H. Cho, COPDGene Investigators, ECLIPSE Investigators, LifeLines Investigators, SPIROMICS Research Group, International COPD Genetics Network Investigators, UK BiLEVE Investigators, International COPD Genetics Consortium. Nat Genet. 2017 Mar;49(3):426-432.

Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis

Brian D. Hobbs[1,2], Kim de Jong[3,4], **Maxime Lamontagne**[5], Yohan Bossé[5], Nick Shrine[6], María Soler Artigas[6], Louise V. Wain[6], Ian P. Hall[7], Victoria E. Jackson[6,8], Annah B. Wyss[9], Stephanie J. London[10], Kari E. North[11], Nora Franceschini[11], David P. Strachan[12], Terri H. Beaty[13], John E. Hokanson[14], James D. Crapo[15,57], Peter J. Castaldi[1,16], Robert P. Chase[1], Traci M. Bartz[17,18], Susan R. Heckbert[17,19,20], Bruce M. Psaty[17,21,20], Sina A. Gharib[22], Pieter Zanen[23], Jan W. Lammers[24], Matthijs Oudkerk[25], H. J. Groen[26], Nick Locantore[27], Ruth Tal-Singer[27], Stephen I. Rennard[28,29], Wim Timens[30], Peter D. Paré[31], Jeanne C. Latourelle[32], Josée Dupuis[33,34], George T. O'Connor[34,35], Jemma B. Wilk[34], Woo Jin Kim[36], Mi Kyeong Lee[36], Yeon-Mok Oh[37], Judith M. Vonk[3,4], Harry J. de Koning[38], Shuguang Leng[39], Steven A. Belinsky[39], Yohannes Tesfaigzi[39], Ani Manichaikul[40,41], Xin-Qun Wang[41], Stephen S. Rich[40,41], R Graham Barr[42], David Sparrow[43], Augusto L. Litonjua[1,2], Per Bakke[44], Amund Gulsvik[44], Lies Lahousse[45,46], Guy G. Brusselle[47], Bruno H. Stricker[45,48,49,50], André G. Uitterlinden[45,49,50], Elizabeth J. Ampleford[51], Eugene R. Bleecker[51], Prescott G. Woodruff[52], Deborah A. Meyers[51], Dandi Qiao[1], David A. Lomas[53], Jae-Joon Yim[54], Deog Kyeom Kim[55], Iwona Hawrylkiewicz[56], Pawel Sliwinski[56], Megan Hardin[1,2], Tasha E. Fingerlin[57,58], David A. Schwartz[57,59,60], Dirkje S. Postma[26,4], William MacNee[61], Martin D. Tobin[6,62], Edwin K. Silverman[1,2], H. Marike Boezen[3,4], Michael H. Cho[1,2], COPDGene Investigators, ECLIPSE Investigators, LifeLines Investigators, SPIROMICS Research Group, International COPD Genetics Network Investigators, UK BiLEVE Investigators, International COPD Genetics Consortium

1 Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA

2 Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA, USA

3 University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, the Netherlands

4 University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD (GRIAC), Groningen, the Netherlands

5 Institut universitaire de cardiologie et de pneumologie de Québec, Québec, Canada

6 Genetic Epidemiology Group, Department of Health Sciences, University of Leicester, Leicester, UK

7 Division of Respiratory Medicine, Queen's Medical Centre, University of Nottingham, Nottingham, UK

8 Department of Health Sciences, University of Leicester, Leicester, UK

9 Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA

10 Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, NC, USA

11 Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA

12 St George's, University of London, Cranmer Terrace, London SW17 0RE, UK

13 Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA

14 Department of Epidemiology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

15 National Jewish Health, Denver, CO, USA

16 Division of General Internal Medicine, Brigham and Women's Hospital, Boston, MA, USA

17 Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA

18 Departments of Medicine and Biostatistics, University of Washington, Seattle, WA, USA

19 Department of Epidemiology, University of Washington, Seattle, WA, USA

20 Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA

21 Departments of Epidemiology, Medicine and Health Services, University of Washington, Seattle, WA, USA

22 Computational Medicine Core, Center for Lung Biology, UW Medicine Sleep Center, Department of Medicine, University of Washington, Seattle, WA, USA

23 University Medical Center Utrecht, Department of Pulmonary Diseases, Utrecht, the Netherlands

24 Department of Pulmonology, University Medical Center Utrecht, University of Utrecht, Utrecht, the Netherlands

25 Department of Radiology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

26 University of Groningen, University Medical Center Groningen, Department of Pulmonology, Groningen, the Netherlands

27 GSK R&D, King of Prussia, PA, USA

28 Pulmonary, Critical Care, Sleep and Allergy Division, Department of Internal Medicine, University of Nebraska Medical Center, Omaha, NE, USA

29 Clincal Discovery Unit, AstraZeneca, Cambridge, UK

30 Department of Pathology and Medical Biology, University of Groningen, University Medical Center Groningen, GRIAC Research Institute, Groningen, the Netherlands

31 University of British Columbia Center for Heart Lung Innovation and Institute for Heart and Lung Health, St Paul's Hospital, Vancouver, British Columbia, Canada

32 Department of Neurology, Boston University School of Medicine, Boston, MA, USA

33 Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

34 The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA, USA

35 Pulmonary Center, Department of Medicine, Boston University School of Medicine, Boston, MA, USA

36 Department of Internal Medicine and Environmental Health Center, School of Medicine, Kangwon National University, Chuncheon, South Korea

37 Department of Pulmonary and Critical Care Medicine, and Clinical Research Center for Chronic Obstructive Airway Diseases, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea

38 Department of Public Health, Erasmus Medical Center Rotterdam, Rotterdam, the Netherlands

39 Lovelace Respiratory Research Institute, Albuquerque, NM, USA

40 Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA

41 Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA

42 Department of Medicine, College of Physicians and Surgeons and Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA

43 VA Boston Healthcare System and Department of Medicine, Boston University School of Medicine, Boston, MA, USA

44 Department of Clinical Science, University of Bergen, Bergen, Norway

45 Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands

46 Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium

47 Department of Public Health, Erasmus Medical Center, Rotterdam, the Netherlands

48 Netherlands Health Care Inspectorate, The Hague, the Netherlands

49 Department of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands

50 Netherlands Genomics Initiative (NGI)-sponsored Netherlands Consortium for Healthy Aging (NCHA), Leiden, the Netherlands

51 Center for Genomics and Personalized Medicine Research, Wake Forest University School of Medicine, Winston Salem, NC, USA

52 Cardiovascular Research Institute and the Department of Medicine, Division of Pulmonary, Critical Care, Sleep, and Allergy, University of California at San Francisco, San Francisco, CA, USA

53 University College London, London, UK

54 Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Seoul National University College of Medicine, South Korea

55 Seoul National University College of Medicine, SMG-SNU Boramae Medical Center, Seoul, South Korea

56 National Tuberculosis and Lung Disease Research Institute, Warsaw, Poland

57 Center for Genes, Environment and Health, National Jewish Health, Denver, CO, USA

58 Department of Biostatistics and Informatics, University of Colorado

Denver, Aurora, CO, USA

59 Department of Medicine, School of Medicine, University of Colorado Denver, Aurora, CO, USA

60 Department of Immunology, School of Medicine, University of Colorado Denver, Aurora, CO, USA

61 University of Edinburgh, Edinburgh, UK

62 National Institute for Health Research (NIHR) Leicester Respiratory Biomedical Research Unit, Glenfield Hospital, Leicester, UK

*Corresponding author:

Michael H. Cho (remhc@channing.harvard.edu)

tel: 617-525-0897

fax: 888-487-1078

Résumé

La maladie pulmonaire obstructive chronique (MPOC) est l'une des principales causes de mortalité à l'échelle mondiale. Nous avons réalisé une étude pangénomique avec 15,256 cas et 47,936 témoins, puis répliqué les résultats avec un $P < 5x10^{-6}$ dans 9,498 cas et 9,748 témoins supplémentaires. Dans la méta-analyse des deux étapes, 22 régions chromosomiques ont atteint le seuil de significativité, dont 9 précédemment associées à la MPOC, 15 aux fonctions respiratoires, et quatre nouvelles (*EEFSEC*, *DSP*, *MTCL1*, et *SFTPD*). Deux des loci ont précédemment été associés à la fibrose pulmonaire (*FAM13A* et *DSP*), mais les allèles de risque des deux maladies sont dans des directions opposées. Aucune de nos régions ne chevauchent les résultats des GWAS sur l'asthme, mais un locus connu pour le risque commun entre l'asthme et l'obésité fut identifié dans notre analyse. Nous avons également identifié une corrélation génétique entre la MPOC et l'asthme. Cette étude a identifié de nouvelles régions, démontré le lien important entre la MPOC et les fonctions respiratoires, et identifié des loci partagés entre la MPOC et d'autres maladies respiratoires.

Abstract

Chronic obstructive pulmonary disease (COPD) is a leading cause of mortality worldwide[1]. We performed a genetic association study in 15,256 cases and 47,936 controls, with genotyping of select top results ($P < 5x10^{-6}$) in 9,498 cases and 9,748 controls. In the combined meta-analysis, we identified 22 loci at genome-wide significance, of which 15 have been associated with lung function in general population samples, and 4 (*EEFSEC*, *DSP*, *MTCL1*, and *SFTPD*) are novel. We noted 2 loci shared with pulmonary fibrosis (*FAM13A* and *DSP*) but with opposite risk alleles for COPD. None of our loci overlapped with genome-wide associations for asthma; however, one locus has been implicated in the joint susceptibility to asthma and obesity. We also identified genetic correlation between COPD and asthma. Our findings highlight novel loci, demonstrate the importance of specific lung function loci to COPD, and identify potential regions of genetic overlap between COPD and other respiratory diseases.

Letter

Chronic obstructive pulmonary disease (COPD) is a leading cause of mortality worldwide[1]. We performed a genetic association in 15,256 cases and 47,936 controls, with replication of select top results ($P < 5x10^{-6}$) in 9,498 cases and 9,748 controls. In the combined meta-analysis, we identified 22 loci at genome-wide significance, including 13 new associations with COPD. Nine of these 13 loci have been associated with lung function in general population samples[2-7]; however, 4 (*EEFSEC*, *DSP*, *MTCL1*, and *SFTPD*) are novel. We noted 2 loci shared with pulmonary fibrosis[8,9] (*FAM13A* and *DSP*) but with opposite risk alleles for COPD. None of our loci overlapped with genome-wide associations for asthma; however, one locus has been implicated in the joint susceptibility to asthma and obesity[10]. We also identified genetic correlation between COPD and asthma. Our findings highlight novel loci, demonstrate the importance of specific lung function loci to COPD, and identify potential regions of genetic overlap between COPD and other respiratory diseases.

COPD is characterized by persistent and progressive airflow limitation diagnosed by lung function testing[1]. While cigarette smoking is the major risk factor, susceptibility is also influenced by genetics[11-13]. We established the International COPD Genetics Consortium (ICGC) to coordinate efforts to find susceptibility loci[14]. We defined cases based on pre-bronchodilator evidence of moderate-to-severe airflow limitation by modified GOLD criteria[15]; controls had normal spirometry, and all analyses were adjusted for age, sex, and cigarette smoking (pack-years and smoking status). We performed a two-stage genome-wide association study (**Figure 5-1**). In Stage 1, we combined 26 cohorts (**Supplementary Table 5-S1 and 5-S2**) containing 63,192 individuals (15,256 COPD cases and 47,936 controls). We selected 79 loci with $P < 5x10^{-6}$ and in analysis Stage 2, we tested them in the UK BiLEVE dataset (9,498 COPD cases and 9,748 controls) from the UK Biobank and performed an overall meta-analysis (**Supplementary Table 5-S3**).

We identified 13 genome-wide significant ($P < 5x10^{-8}$) associations in Stage 1. Following the Stage 2 analysis, an additional 9 loci achieved genome-wide significance in the overall meta-analysis (**Table 5-1**, **Figure 5-2**, **Supplementary Figures 5-S1 and 5-S2**). Analysis of only European ancestry (**Supplementary Table 5-S4**) and only African

ancestry (**Supplementary Table 5-S5** and **Supplementary Figure 5-S3**) Stage 1 cohorts showed no unique association signals. Of the 22 genome-wide significant loci described in our study, 9 have been previously described as genome- (or exome-) wide significant in studies of COPD[13,16-19]: *HHIP*, *CHRNA5*, *HTR4*, *FAM13A*, *RIN3*, *TGFB2*, *GSTCD-NPNT*, *CYP2A6*, and *IL27-CCDC101*. The remaining 13 loci have not been previously associated with COPD at genome- (or exome-) wide significance. Eight of these 13 loci: *ADGRG6*/*GPR126*, *THSD4*, *ADAM19*, *TET2*, *CFDP1*, *AGER*, *ARMC2*, and *RARB* have been previously described and replicated (**Supplementary Table 5-S6**) in general population GWASs of two measures of lung function (FEV_1 and FEV_1/FVC) that are used in conjunction to diagnose COPD[2,4-7,20,21]. One locus near *PID1* was previously associated with FEV_1/FVC, but had not replicated in those studies[4,6]. Four loci are newly being described as genome-wide significant in association with either COPD or lung function: *EEFSEC*, *DSP*, *MTCL1*, and *SFTPD* (**Figure 5-3**).

To explore the potential function and causal genes for our novel loci, in addition to using publicly available datasets and prioritization tools (**Supplementary Table 5-S7**), we also examined a larger set of lung expression quantitative trait loci (eQTL) in 1038 subjects, including subjects with COPD[22] (**Supplementary Table 5-S8**). As eQTL are pervasive, we also attempted to determine whether our association signal co-localized[23] with an eQTL signal in lung tissue (**Supplementary Table 5-S9**). We found strong evidence of co-localization (posterior probability > 0.8) for *DSP*, a major protein of desmosomes required for epidermal integrity [24], and *MTCL1*, important in epithelial-cell-specific microtubule stabilization[25,26], and expressed in respiratory epithelial cells[27]. Variants in strong LD with our top *MTCL1* variant rs647097 (NC_000018.9:g.8808464T>C) appear to have enhancer histone marks in fetal lung fibroblasts[28,29]. In contrast, we found no evidence of a strong eQTL signal or co-localization at our other two novel loci. At 3q21, *EEFSEC* is a potential candidate, as it is a paralog of *TUFM*, a top blood and lung eQTL gene for the 16p11.2/*IL27* COPD susceptibility locus[19], recently part of a novel COPD-related pathway involving *NLRX1*[30-32]. At 10q22, pulmonary surfactant-associated protein D (*SFTPD*) is the most likely candidate, as it is highly expressed in pneumocytes[27], and *sftpd* (-/-) mice develop pulmonary emphysema[33]. SFTPD has been explored as a COPD biomarker[34], and while

rs721917 (NC_000010.10:g.81706324A>G) is not an eQTL, polymorphisms in *SFTPD*, including rs721917, may lead to decreased surfactant protein D levels[35]; though the association of SFTPD polymorphisms with COPD susceptibility have been inconsistent. Our analysis also led to some additional insights into other previously described loci. We found evidence of COPD association and eQTL statistical co-localization in lung tissue (posterior probability > 0.8) for *THSD4*, *HHIP*, *AGER*, *CHRNA3*, and *RARB* (**Supplementary Table 5-S9**). Additional data on eQTLs (**Supplementary Table 5-S8**), cohort-specific associations at each locus (**Supplementary Figures S1a-v**), fine mapping (Supplementary Note and **Supplementary Table 5-S10**), causal gene (**Supplementary Table 5-S11 and 5-S12**), and other supportive analysis for previously described and novel loci can be found in the Supplementary Note.

We note that our top variant at *DSP* (rs2076295, NC_000006.11:g.7563232T>G) is also associated ($P = 1.1x10^{-19}$) with pulmonary fibrosis[8]. Recently, a re-sequencing study[36] at the *DSP* locus identified a second fibrosis-associated variant, rs2744371 (NC_000006.11:g.7554174A>C) with $P_{fibrosis} = 0.002$ and $P_{COPD} = 0.04$. We also note overlap at the *FAM13A* locus (top fibrosis SNP[8], rs2609255 [NC_000004.11:g.89811195G>T]; $P_{fibrosis} = 2.2x10^{-11}$, $P_{COPD} = 1.9x10^{-7}$). We performed additional analysis to investigate genetic overlap using gwas-pw[37] (see Supplementary Note). We confirmed overlap at the *DSP* and *FAM13A* loci with a posterior probability of > 0.99, and additionally discovered overlap near *MAPT*/*KANSL1* (top fibrosis SNP[8], rs1981997 [NC_000017.10:g.44056767G>A]; $P_{fibrosis} = 8.87\times10^{-14}$, $P_{COPD} = 4.5x10^{-3}$) with posterior probability of 0.84. While the *MAPT*/*KANSL1* locus did not reach genome-wide significance in our study, we note its independent discovery in a genome-wide association in extremes of lung function[7]. Notably, for all four of these variants (in *DSP* [2], *FAM13A*, and *MAPT*), the fibrosis risk allele is protective for development of COPD. Emphysema, a key component of COPD, and pulmonary fibrosis are both smoking-related lung diseases that have both shared and distinct pathophysiology[38-40], though genetic loci with opposing effects have not been previously described. Additional investigation of these loci as a well as a more comprehensive assessment of genetic overlap of COPD and pulmonary fibrosis may lead to insight into both disorders.

Because our analysis relied on a spirometric definition of COPD alone, we did not specifically exclude other causes of airway obstruction such as asthma, which can overlap with COPD in adults[41]. To define COPD, we used pre-bronchodilator spirometry, which was available across all cohorts, and we included at least moderately affected cases (FEV_1 < 80% predicted). We examined the top set of genome-wide significant results in a subset of our largest cohorts with both pre- and post-bronchodilator data and densely imputed genotypes; overall, the effect sizes (mean difference = 0.001) and P values (mean $\log_{10}$ P value difference = 0.18) were similar (**Supplementary Table 5-S13** and **Supplementary Figures 5-S4 and 5-S5**). In addition, a recent GWAS of FEV_1, FVC, and FEV_1/FVC did not find substantial differences including and excluding subjects with asthma[7]. In the 79 variants tested in Stage 2, we found no significant difference in the OR for COPD association when including and excluding individuals with asthma (**Supplementary Figure 5-S6**).

We examined COPD associations of genome-wide significant asthma (and asthma-associated traits) loci from the NHGRI-EBI GWAS Catalog[42] (**Supplementary Table 5-S14**). We also compared our COPD association results to the GABRIEL asthma study[43] (**Supplementary Tables 5-S15**). None of the genome-wide significant loci from asthma and COPD overlapped. Further, no asthma or COPD loci showed Bonferroni-adjusted (for number of look-ups) significant association with the other disease, though several loci showed nominal (P<0.05) significance. The 16p11.2 (*CCDC101*) locus has been described in the joint susceptibility to asthma and obesity[10]. COPD susceptibility is strongly related to cigarette smoking. Two of our loci (15q25 and 19q13) have been previously associated with smoking behavior[44,45], though we found no additional evidence of overlap in genome-wide significant variants described in the NHGRI-EBI GWAS Catalog[42] and Tobacco and Genetics Consortium GWAS[45] (**Supplementary Tables 5-S16, 5-S17 and 5-S18**). We additionally evaluated overlap of our top 22 loci with COPD comorbidities (**Supplementary Table 5-S19**) and radiographic imaging features **(Supplementary Table 5-S20**).

In contrast to minimal overlap in genome-wide significant results with asthma and smoking, we discovered a significant overall genetic correlation of COPD with asthma

($r_{genetic}$ = 0.38, P = $6.2x10^{-5}$) using LD score regression in our European-ancestry subjects[46,47]. We also assessed genetic correlation with population-based lung function, pulmonary fibrosis, smoking behavior, and two common COPD comorbidities, coronary artery disease and osteoporosis. We identified significant correlation of COPD with lung function and two aspects of smoking behavior, but not with common comorbidities or with pulmonary fibrosis (**Figure 5-4**). The lack of significant correlation of COPD with pulmonary fibrosis may indicate our overlapping loci for COPD and pulmonary fibrosis are not representative of a broader disease correlation; alternatively, it could reflect limited sample size or a mix of positive and negative genetic correlations across the genome for the diseases. In potential support of this latter hypothesis, and in contrast to the loci we describe in this study, are recent descriptions of rare variants in telomerase genes predisposing to both emphysema, a key feature of COPD, and pulmonary fibrosis[40,48]. Our analysis of partitioned heritability identified COPD genetic association enrichment in fetal lung tissue (coefficient P = $3.5x10^{-7}$); other bioinformatics analyses also support functional annotation of COPD associations to lung tissue or lung cell types (Supplementary Note).

Our study is, to our knowledge, the largest genome-wide association study of COPD cases to date and includes over 60,000 subjects (including 15,256 COPD cases) in the Stage 1 analysis. We chose to combine subjects of different ethnicities, hypothesizing that shared COPD risk factors across ethnicities would outweigh power loss due to heterogeneity. While methods have been developed that can more rigorously assess the degree of overlap and provide additional power in this setting[49], none of our non-white cohorts were sufficiently sized or powered for these analyses. COPD is also a highly heterogeneous disease; whether a more precise phenotypic definition would result in greater power is not clear. We used a staged study design and examined overall meta-analysis P-values to determine genome-wide significance. Thus, 9 loci (*TET2*, *CFDP1*, *TGFB2*, *AGER*, *ARMC2*, *PID1*, *MTCL1*, *SFTPD*, and *CYP2A6*) from our Stage 1 analysis, which only reached genome-wide significance in either the Stage 2 UK BiLEVE analysis or the overall meta-analysis, should be further replicated. However, six of these 9 association signals are significant if we consider a Bonferroni correction (P < $6.3x10^{-4}$) for the 79 variants tested in Stage 2. Further, 8 of these 9 variants are more strongly associated in the overall meta-

analysis compared to Stage 1; the exception is *RARB*, which has a previously reported association with both lung function[4] and airflow obstruction[21] (**Table 5-1**).

The majority of our significant loci overlap with lung function loci, strengthening the foundation for investigating the relationship of lung function variability in the general population to risk of developing COPD. These loci are unlikely to reflect susceptibility for asthma or for cigarette smoking; however, our association results as a whole show evidence of shared heritability with asthma (supporting investigation into shared genetic etiologies for these diseases) and cigarette smoking behavior (despite adjustment for smoking in our statistical model). We identified functional annotation enrichment for fetal lung cells, supporting a role for early life events contributing to future risk of COPD. Finally, we identify loci that overlap with pulmonary fibrosis, but with opposite risk alleles. Our study highlights the important contribution of genetic association studies to understanding COPD, not only by identifying novel loci, but also illustrating relationships with other pulmonary traits and diseases.

Data Availability Statement

The genome-wide association summary statistics generated in the Stage 1 analysis of the current study are available in the dbGaP repository, https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000179.v5.p2

The Stage 2 analysis summary statistics are available in Supplementary Table 3.

Author Contributions

B.D.H. & M.H.C. contributed to the study concept and design, data analysis, statistical support, and manuscript writing. K.d.J., A.B.W., S.J.L., & D.P.S. contributed to the study concept and design & data analysis. N.S. & M.S.A. contributed to the data analysis and statistical support. T.H.B. & J.E.H. contributed to the study concept and design and statistical support. L.L. contributed to the data collection, data analysis, and statistical support. K.E.N. contributed to data collection and data analysis. J.D.C., B.M.P., N.L., R.T.S., G.T.O., Y.T., R.G.B., S.I.R., P.B., A.G., P.G.W., D.A.M., D.A.S, & E.K.S. contributed to the study concept and design and to data collection. D.Q., T.A.F., M.L., Y.B., N.S., N.F., P.J.C., R.P.C., T.M.B., S.A.G., J.C.L., J.D., J.B.W., M.K.L., S.L., A.M., X.W., & E.J.A. contributed to the data analysis. L.V.W., I.P.H., P.D.P., D.S.P., W.M., M.D.T., & H.M.B. contributed to the study concept and design. S.R.H., M.O., J.V., P.A.D., W.J.K., Y.O., S.S.R., D.S., A.A.L., G.G.B., B.H.S., A.G.U., E.R.B., D.A.L., J.J.Y., D.K.K., I.H., P.S. & M.H. contributed to data collection. All authors, including those whose initials are not listed above, contributed to the critical review and editing of the manuscript and approved the final version of the manuscript.

Competing Financial Interest Statements :

I.P.H. has received grant support from Pfizer. P.J.C. has received research funding from GSK. B.P. serves on the DSMB of a clinical trial funded by the manufacturer and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. N.L. and R.T-S. are shareholders and employees of GSK. S.I.R. is a current employee and shareholder at AstraZeneca. He has served as a consultant, participated in advisory boards, received honorarium for speaking or grant support from: American Board of Internal Medicine, Advantage Healthcare, Almirall, American Thoracic Society, AstraZeneca, Baxter, Boehringer Ingelheim, Chiesi, ClearView Healthcare, Cleveland Clinic, Complete Medical Group, CSL, Dailchi Sankyo, Decision Resources, Forest, Gerson Lehman, Grifols, GroupH, Guidepoint Global, Haymarket, Huron Consulting, Inthought, Johnson and Johnson, Methodist Health System – Dallas, NCI Consulting, Novartis, Pearl, Penn Technology, Pfizer, PlanningShop, PSL FirstWord, Qwessential, Takeda, Theron and WebMD. W.T. reports reports fees to the Department, all outside the submitted work, from Pfizer, GSK, Chiesi, Roche Diagnostics / Ventana, Biotest, Merck Sharp Dohme, Novartis, Lilly Oncology, Boehringer Ingelheim, and grants from Dutch Asthma Fund. J.C.L. is currently an employee of GNS Healthcare in Cambridge, MA. J.B.W. was employed by Pfizer during the time this research was performed. P.B. has received consulting and lecture fees from AstraZeneca, Boehringer Ingelheim, Chiesi, Novartis, and Teva. L.L. has performed consultancy for Boehringer Ingelheim GmbH, received an AstraZeneca Scientific Award and travel support from Novartis, European Respiratory Society, and the Belgian Respiratory Society. P.G.W. has consulted for Amgen, Sanofi, Novartis, Genentech/Roche, Boehringer-Ingelheim, Neostem and has had research grants from Pfizer and Genentech. D.L. received grant support, honoraria and consultancy fees from GSK for work on the ICGN and ECLIPSE studies, and was a member and then Chaired the GSK Respiratory Therapy Area Board (2009-2015). M.H. is a current employee at AstraZeneca. D.A.S. is serving on the scientific advisory boards of Apellis Pharmaceuticals and Pliant Therapeutics, and is the founder and owner of Eleven P15. D.S.P. - The University of Groningen has received money for Professor Postma regarding a grant for research from Astra Zeneca, Chiesi, Genentec, GSK and Roche. Fees for

consultancies were given to the University of Groningen by Astra Zeneca, Boehringer Ingelheim, Chiesi, GSK, Takeda and TEVA. E.K.S. has received honoraria and consulting fees from Merck, grant support and consulting fees from GSK, and honoraria and travel support from Novartis. M.H.C. has received grant support from GSK.

References

1. Vestbo, J. et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med 187, 347-65 (2013).

2. Hancock, D.B. et al. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. Nat Genet 42, 45-52 (2010).

3. Repapi, E. et al. Genome-wide association study identifies five loci associated with lung function. Nat Genet 42, 36-44 (2010).

4. Soler Artigas, M. et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. Nat Genet 43, 1082-90 (2011).

5. Hancock, D.B. et al. Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. PLoS Genet 8, e1003098 (2012).

6. Soler Artigas, M. et al. Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. Nat Commun 6, 8658 (2015).

7. Wain, L.V. et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. Lancet Respir Med 3, 769-81 (2015).

8. Fingerlin, T.E. et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. Nat Genet 45, 613-20 (2013).

9. Fingerlin, T.E. et al. Genome-wide imputation study identifies novel HLA locus for pulmonary fibrosis and potential role for auto-immunity in fibrotic idiopathic interstitial pneumonia. BMC Genet 17, 74 (2016).

10. Gonzalez, J.R. et al. A common 16p11.2 inversion underlies the joint susceptibility to asthma and obesity. Am J Hum Genet 94, 361-72 (2014).

11. Laurell, C.-B. & Eriksson, S. The electrophoretic alpha-1-globulin pattern of serum in alpha-1-antitrypsin deficiency. Scandinavian Journal of Clinical and Laboratory Investigation 15, 132-140 (1963).

12. Silverman, E.K. et al. Genome-wide linkage analysis of severe, early-onset chronic obstructive pulmonary disease: airflow obstruction and chronic bronchitis phenotypes. Hum Mol Genet 11, 623-32 (2002).

13. Cho, M.H. et al. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. Lancet Respir Med 2, 214-25 (2014).

14. Silverman, E.K. et al. Opportunities and challenges in the genetics of COPD 2010: an International COPD Genetics Conference report. COPD 8, 121-35 (2011).

15. Mannino, D.M. & Buist, A.S. Global burden of COPD: risk factors, prevalence, and future trends. Lancet 370, 765-73 (2007).

16. Pillai, S.G. et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. PLoS Genet 5, e1000421 (2009).

17. Cho, M.H. et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. Nat Genet 42, 200-2 (2010).

18. Cho, M.H. et al. A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. Hum Mol Genet 21, 947-57 (2012).

19. Hobbs, B.D. et al. Exome Array Analysis Identifies a Common Variant in IL27 Associated with Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med 194, 48-57 (2016).

20. Wilk, J.B. et al. A genome-wide association study of pulmonary function measures in the Framingham Heart Study. PLoS Genet 5, e1000429 (2009).

21. Wilk, J.B. et al. Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. Am J Respir Crit Care Med 186, 622-32 (2012).

22. Hao, K. et al. Lung eQTLs to help reveal the molecular underpinnings of asthma. PLoS Genet 8, e1003029 (2012).

23. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 10, e1004383 (2014).

24. Vasioukhin, V., Bowers, E., Bauer, C., Degenstein, L. & Fuchs, E. Desmoplakin is essential in epidermal sheet formation. Nat Cell Biol 3, 1076-85 (2001).

25. Sato, Y. et al. The novel PAR-1-binding protein MTCL1 has crucial roles in organizing microtubules in polarizing epithelial cells. J Cell Sci 126, 4671-83 (2013).

26. Sato, Y. et al. MTCL1 crosslinks and stabilizes non-centrosomal microtubules on the Golgi membrane. Nat Commun 5, 5266 (2014).

27. Uhlen, M. et al. Proteomics. Tissue-based map of the human proteome. Science 347, 1260419 (2015).

28. Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 40, D930-4 (2012).

29. Ward, L.D. & Kellis, M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res (2015).

30. Lei, Y. et al. The mitochondrial proteins NLRX1 and TUFM form a complex that regulates type I interferon and autophagy. Immunity 36, 933-46 (2012).

31. Lei, Y., Wen, H. & Ting, J.P. The NLR protein, NLRX1, and its partner, TUFM, reduce type I interferon, and enhance autophagy. Autophagy 9, 432-3 (2013).

32. Kang, M.J. et al. Suppression of NLRX1 in chronic obstructive pulmonary disease. J Clin Invest 125, 2458-62 (2015).

33. Wert, S.E. et al. Increased metalloproteinase activity, oxidant production, and emphysema in surfactant protein D gene-inactivated mice. Proc Natl Acad Sci U S A 97, 5972-7 (2000).

34. Lomas, D.A. et al. Serum surfactant protein D is steroid sensitive and associated with exacerbations of COPD. Eur Respir J 34, 95-102 (2009).

35. Foreman, M.G. et al. Polymorphisms in surfactant protein-D are associated with chronic obstructive pulmonary disease. Am J Respir Cell Mol Biol 44, 316-22 (2011).

36. Mathai, S.K. et al. Desmoplakin Variants Are Associated with Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med 193, 1151-60 (2016).

37. Pickrell, J.K. et al. Detection and interpretation of shared genetic influences on 42 human traits. Nat Genet 48, 709-17 (2016).

38. Washko, G.R. et al. Lung volumes and emphysema in smokers with interstitial lung abnormalities. N Engl J Med 364, 897-906 (2011).

39. Chilosi, M., Poletti, V. & Rossi, A. The pathogenesis of COPD and IPF: distinct horns of the same devil? Respir Res 13, 3 (2012).

40. Stanley, S.E. et al. Telomerase mutations in smokers with severe emphysema. J Clin Invest 125, 563-70 (2015).

41. Soriano, J.B. et al. The proportional Venn diagram of obstructive lung disease: two approximations from the United States and the United Kingdom. Chest 124, 474-81 (2003).

42. Welter, D. et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 42, D1001-6 (2014).

43. Moffatt, M.F. et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med 363, 1211-21 (2010).

44. Thorgeirsson, T.E. et al. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. Nat Genet 42, 448-53 (2010).

45. Consortium, T.a.G. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat Genet 42, 441-7 (2010).

46. Finucane, H.K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet 47, 1228-35 (2015).

47. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. Nat Genet 47, 1236-41 (2015).

48. Stanley, S.E. et al. Loss-of-function mutations in the RNA biogenesis factor NAF1 predispose to pulmonary fibrosis-emphysema. Sci Transl Med 8, 351ra107 (2016).

49. Coram, M.A. et al. Leveraging Multi-ethnic Evidence for Mapping Complex Traits in Minority Populations: An Empirical Bayes Approach. Am J Hum Genet 96, 740-52 (2015).

50. Consortium, C.A.D. et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet 45, 25-33 (2013).

51. Wood, A.R. et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet 46, 1173-86 (2014).

52. Locke, A.E. et al. Genetic studies of body mass index yield new insights for obesity biology. Nature 518, 197-206 (2015).

53. Zheng, H.F. et al. Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. Nature 526, 112-7 (2015).

Table

Table 5-1 : Overall study results showing 22 loci with genome-wide significant P values in overall meta-analysis following UK BiLEVE Stage 2 analysis.

rsID	Closest Gene	Locus	Risk Allele	Alt Allele	Risk Allele Mean Freq (Range)	Stage 1 OR (95% CI)	Stage 1 P Value	UK BiLEVE (Stage 2) OR (95% CI)	UK BiLEVE (Stage 2) P Value	**Overall Meta-Analysis P Value**
rs13141641	*HHIP*	4q31.21	T	C	0.594 (0.524-0.886)	1.231 (1.183-1.281)	1.16E-24	1.213 (1.16-1.267)	8.15E-18	**9.10E-41**
rs17486278	*CHRNA5*	15q25.1	C	A	0.351 (0.244-0.442)	1.224 (1.177-1.273)	2.61E-24	1.126 (1.077-1.178)	2.35E-07	**1.77E-28**
rs7733088	*HTR4*	5q32	G	A	0.602 (0.47-0.685)	1.178 (1.129-1.229)	4.40E-14	1.177 (1.127-1.229)	1.78E-13	**5.33E-26**
rs9399401	*ADGRG6*	6q24.1	T	C	0.724 (0.615-0.748)	1.141 (1.095-1.189)	3.59E-10	1.174 (1.119-1.232)	6.18E-11	**1.81E-19**
rs1441358	*THSD4*	15q23	G	T	0.332 (0.194-0.546)	1.132 (1.09-1.177)	2.06E-10	1.121 (1.071-1.172)	6.87E-07	**8.22E-16**
rs6837671	*FAM13A*	4q22.1	G	A	0.405 (0.364-0.582)	1.157 (1.115-1.2)	1.02E-14	1.066 (1.021-1.114)	3.75E-03	**7.48E-15**
rs11727735	*GSTCD*	4q24	A	G	0.936 (0.926-0.988)	1.266 (1.167-1.374)	1.55E-08	1.246 (1.144-1.357)	4.93E-07	**3.84E-14**
rs754388	*RIN3*	14q32.12	C	G	0.821 (0.804-0.865)	1.196 (1.136-1.258)	7.07E-12	1.108 (1.05-1.169)	1.85E-04	**4.96E-14**
rs113897301	*ADAM19*	5q33.3	AT	A	0.175 (0.052-0.187)	1.195 (1.13-1.263)	4.52E-10	1.127 (1.066-1.192)	2.79E-05	**1.58E-13**
rs2047409	*TET2*	4q24	A	G	0.618 (0.222-0.649)	1.101 (1.059-1.146)	1.58E-06	1.136 (1.087-1.188)	1.95E-08	**2.46E-13*†**
rs2955083	*EEFSEC*	3q21.3	A	T	0.881 (0.854-0.893)	1.195 (1.123-1.272)	2.00E-08	1.167 (1.093-1.246)	4.01E-06	**4.16E-13**
rs7186831	*CFDP1*	16q23.1	A	G	0.435 (0.226-0.472)	1.125 (1.07-1.183)	3.54E-06	1.116 (1.069-1.165)	6.63E-07	**1.12E-11*†**
rs10429950	*TGFB2*	1q41	T	C	0.731 (0.216-0.773)	1.117 (1.071-1.164)	1.83E-07	1.096 (1.044-1.15)	1.94E-04	**1.66E-10*†**
rs2070600	*AGER*	6p21.32	C	T	0.955 (0.85-0.987)	1.276 (1.151-1.414)	3.54E-06	1.208 (1.105-1.319)	2.96E-05	**5.94E-10*†**
rs17707300	*CCDC101*	16p11.2	C	T	0.373 (0.109-0.433)	1.122 (1.079-1.167)	6.24E-09	1.063 (1.018-1.111)	6.10E-03	**6.75E-10**
rs2806356	*ARMC2*	6q21	C	T	0.184 (0.052-0.242)	1.125 (1.071-1.181)	2.84E-06	1.116 (1.057-1.178)	6.88E-05	**8.34E-10*†**
rs16825267	*PID1*	2q36.3	C	G	0.929 (0.874-0.942)	1.242 (1.149-1.342)	5.22E-08	1.131 (1.045-1.224)	2.27E-03	**1.68E-09***
rs2076295	*DSP*	6p24.3	T	G	0.554 (0.442-0.581)	1.107 (1.067-1.149)	4.95E-08	1.061 (1.016-1.107)	7.45E-03	**3.97E-09**
rs647097	*MTCL1*	18p11.22	C	T	0.269 (0.259-0.399)	1.106 (1.06-1.154)	3.03E-06	1.089 (1.038-1.142)	4.66E-04	**6.14E-09*†**
rs1529672	*RARB*	3p24.2	C	A	0.831 (0.675-0.862)	1.162 (1.106-1.221)	2.37E-09	1.048 (0.991-1.109)	9.95E-02	**2.47E-08**
rs721917	*SFTPD*	10q22.3	G	A	0.422 (0.392-0.631)	1.092 (1.053-1.132)	2.11E-06	1.068 (1.023-1.115)	2.60E-03	**2.49E-08***
rs12459249	*CYP2A6*	19q13.2	C	T	0.662 (0.617-0.702)	1.126 (1.071-1.183)	2.89E-06	1.077 (1.029-1.127)	1.35E-03	**3.42E-08***

OR = odds ratio, CI = confidence interval. *Genome-wide significant in overall analysis only †Stage 2 results significant after Bonferroni.

Figures

Stage 1:
26 studies represented

Study	COPD cases	Controls
ARIC	1,060	6,164
B58	205	3,665
CHS EA	736	1,586
COPACETIC	397	1,906
COPDGene NHW	3,068	2,110
ECLIPSE	1,741	149
EOCOPD*	394	495
ICGN*	1,852	557
EQTL	252	224
FHS	701	5,110
LifeLines	466	9,863
Lovelace	259	641
MESA Caucasian	167	754
NETT-NAS	376	435
Norway/GenKOLS	846	695
RS1	112	815
RS2	94	811
RS3	106	1,596
SPIROMICS	571	175
TCGS-Poland*	307	311
CHS AA	138	292
COPDGene AA	910	1,556
KARE	199	6,741
MESA AA	94	532
MESA hispanic	52	548
TCGS-Korea*	153	205
Total	15,256	47,936

Stage 2:
Top results tested in UK BiLEVE study

Study	COPD cases	Controls
UK BiLEVE never smokers	3,737	4,871
UK BiLEVE heavy smokers	5,761	4,877
Total	9,498	9,748

Figure 5-1 : Study design showing the cohorts used in each stage of the analysis.

ARIC = Atherosclerosis Risk in Communities Study, B58 = British 1958 Birth Cohort, CHS = Cardiovascular Health Study, COPACETIC = COPD Pathology: Addressing Critical gaps, Early Treatment & Diagnosis and Innovative Concepts, ECLIPSE = Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points, eQTL = Lung Expression Quantitative Trait Loci Study, FHS = Framingham Heart Study, KARE = Korean Association Resource project, MESA = Multi-Ethnic Study of Atherosclerosis, NETT-NAS = National Emphysema Treatment Trial / Normative Aging Study, RS = Rotterdam Study, SPIROMICS = Subpopulations and intermediate outcome measures in COPD study , EOCOPD = Boston Early-Onset COPD Study, ICGN = International COPD Genetics Network, TCGS = Transcontinental COPD Genetics Study, UK BiLEVE = UK Biobank Lung Exome Variant Evaluation; NHW = Non-Hispanic white, AA = African American, EA = European American.* Studies without genome-wide array genotyping (custom genotyping).

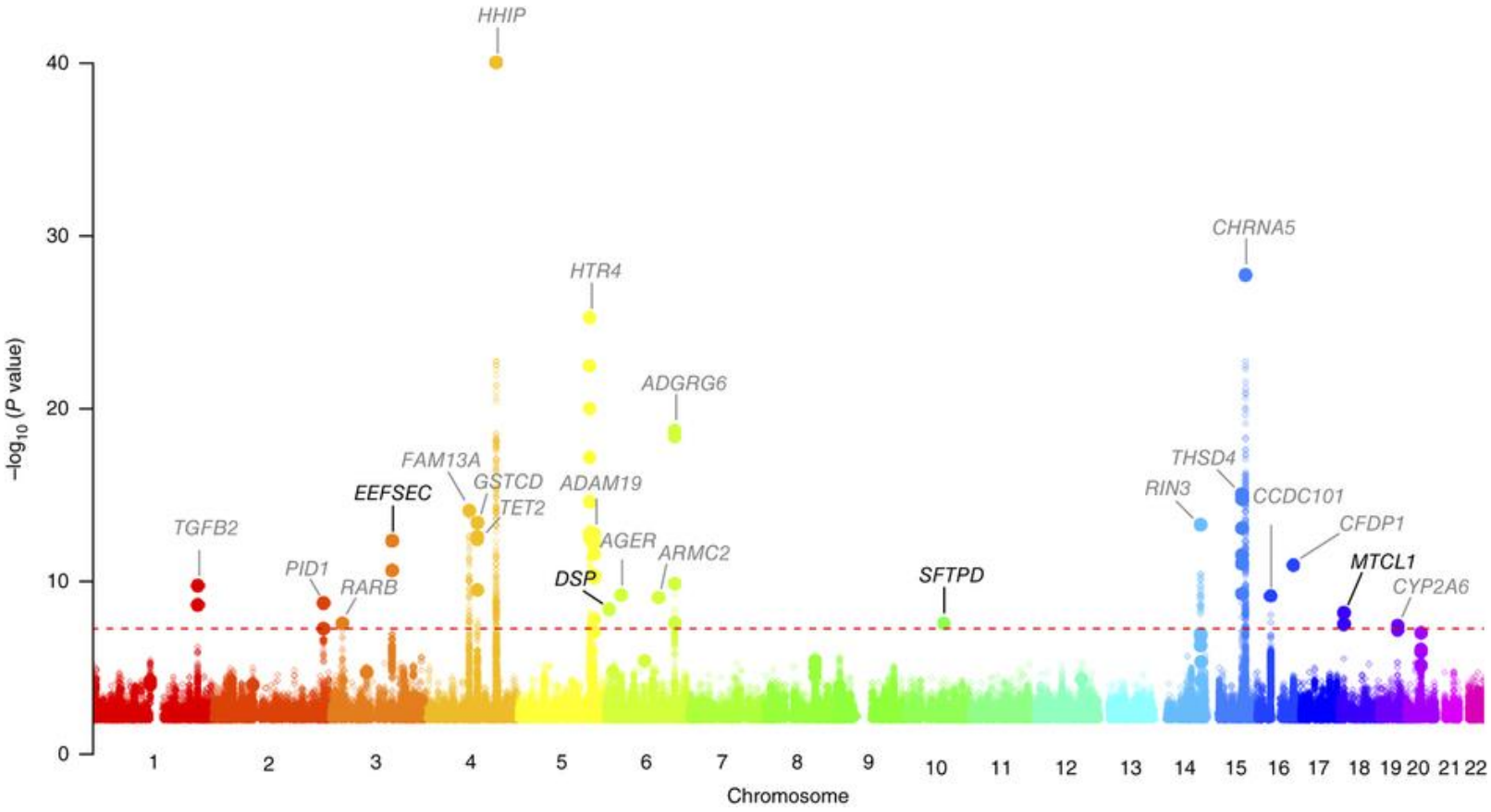

Figure 5-2 : Manhattan plot.

Manhattan plot showing P values for Stage 1 analysis (small open diamonds) with overlay of overall meta-analysis P values for SNPs analyzed in UK BiLEVE Stage 2 analysis (filled circles). Gene names in gray are previously described COPD or lung function (FEV1 or FEV1/FVC) loci; black are novel loci discovered in this study. The Stage 1 cohorts with available genotyping data (Supplementary Figures 1a-v) and the UK BiLEVE cohort determined the sample size for each top variant. The red dashed line indicates the threshold for genome-wide significance (P value < 5x10-8).

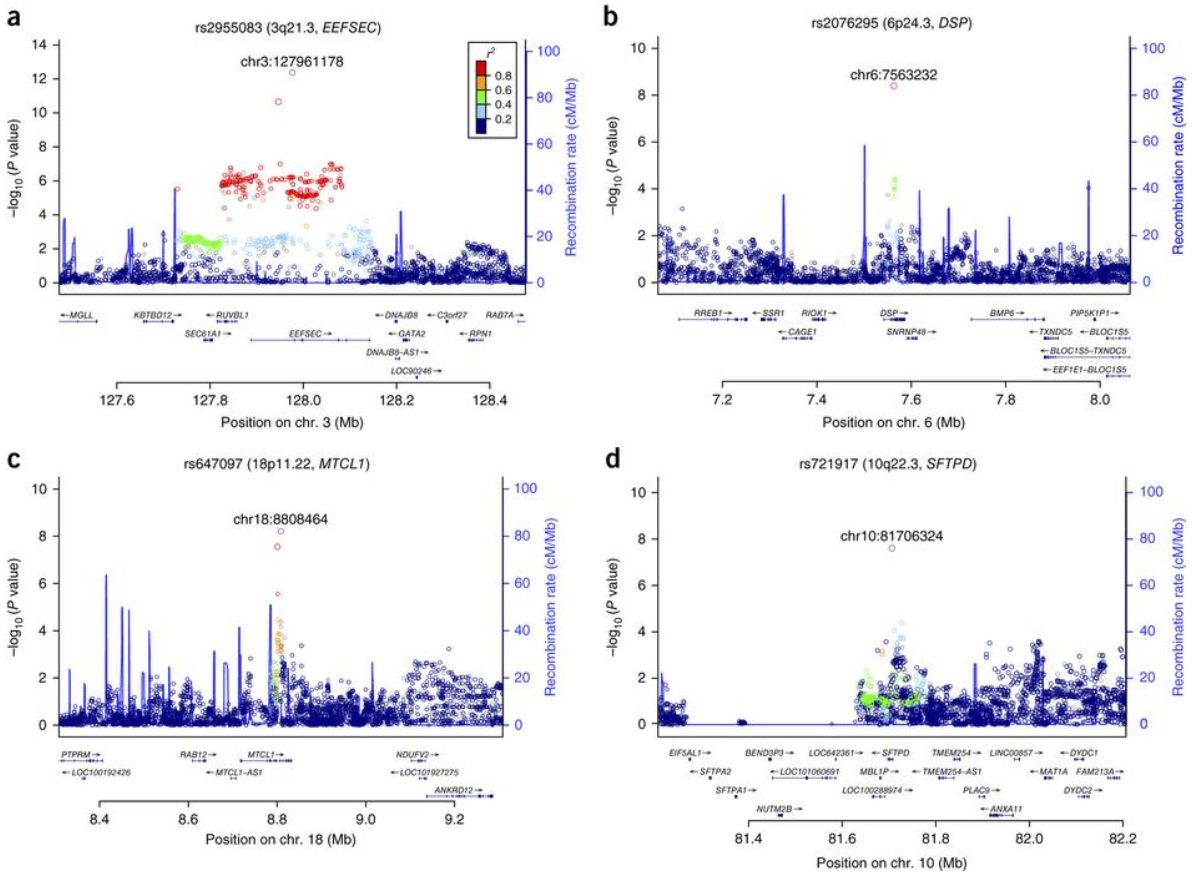

Figure 5-3 : Regional association for novel loci.

LocusZoom plots showing regional association of variants at the four novel COPD loci. The point size is proportional to the sample size, where Stage 1 cohorts with available genotyping data (Supplementary Figures 1a-v) and the UK BiLEVE cohort determined the sample size for each top variant.

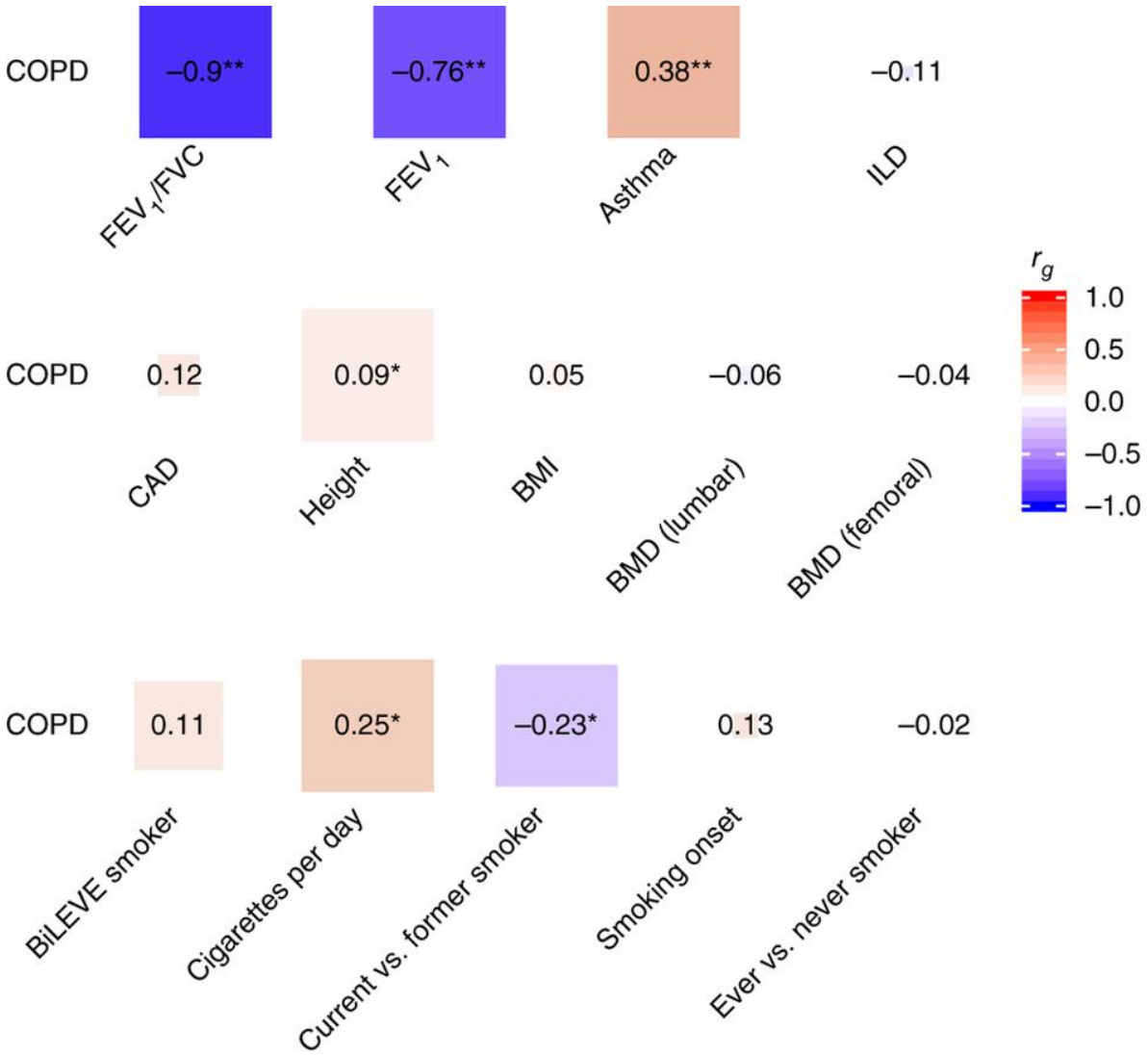

Figure 5-4 : Genetic correlation (using LD score regression) between COPD and other traits.

Shading and numbers represents strength of correlation. An asterisk indicates nominal ($P < 0.05$) significance, and a double asterisk indicates significant after Bonferroni correction for number of pairwise comparisons. fev1fvc and fev1 = lung function (FEV1/FVC ratio and FEV1 from CHARGE/SpiroMeta4, asthma taken from the asthma GWAS by the GABRIEL Consortium43, ild = pulmonary fibrosis from Fingerlin et al.8,9, bilSmk = subset of smokers in the UK BiLEVE study7, smkCpd = cigarettes per day smoking from the Tobacco and Genetics (TAG) Consortium45, smkFormer = current versus former smokers from TAG, smkOnset = age of smoking initiation from TAG, smkEver = ever versus never smoking from TAG. cad = coronary artery disease from the CARDIoGRAM study50, height51 and bmi (body mass index)52 from the GIANT consortium, bmdLumbar and bmdFemoral = lumbar and femoral bone mineral density, respectively, from the Genetic Factors for Osteoporosis (GeFOS) Consortium53.

Supplement

Online Methods

Study Cohorts

We invited investigators from 22 studies with genome-wide association data and COPD case-control or general population samples with spirometry to participate in a genome-wide association meta-analysis. Additionally, we included four cohorts with Illumina HumanExome v1.2 and custom genotyping based primarily on prior top results from a previously published COPD GWAS[13], using results with $P < 1x10^{-4}$ using plink '--clump' on the COPDGene non-Hispanic whites to perform linkage disequilibrium pruning ($r^2 < 0.8$), preferentially retaining both an imputed and genotyped top SNP at each locus. An additional group of variants was a candidate panel, based on results from a previous candidate gene analysis[54], as well as variants identified in association with lung function (supplementing the existing content on the array, which included variants from previous genome-wide association studies), including the lead SNP and a 200kb region around that SNP pruned for variants with $P < 0.01$ and $r^2 < 0.8$, and additional top-ranked SNPs for COPDGene-specific analyses for lung function, bronchodilator responsiveness, exacerbations, and SNPs from candidate genes.

The baseline characteristics of these 26 cohorts can be seen in **Supplementary Table 5-S1**. Each cohort obtained approval from appropriate ethical/regulatory bodies; informed consent was obtained for all individuals. (Further cohort-specific methods can be found in the Data Supplement.) As most of these cohorts did not have post-bronchodilator spirometry, we used a modified definition of GOLD criteria based on pre-bronchodilator spirometry: forced expiratory volume in 1 second (FEV_1) < 80% and FEV_1 to forced vital capacity (FVC) ratio of < 0.7 for cases, and FEV_1 > 80% and $FEV_1/FVC > 0.7$ for controls. Logistic regression was performed in each cohort, adjusting for age, sex, pack-years of smoking, ever-smoking status, current-smoking status, and ancestry-based principal components, as appropriate for each study. Summary statistics were assessed using

EasyQC[55] version 10.1. More detailed cohort information, including cohort-specific methods, can be found in the Supplementary Note.

Genome-wide association quality control

Summary statistics, including effect allele and other allele oriented to the + strand, effect allele frequency, chromosome and position (hg19), and imputation quality were uploaded to a secure site at the Brigham and Women's Hospital / Channing Division of Network Medicine. Quality control assessments included assessing allele frequencies versus 1000 Genomes reference, standard error versus sample size, and quantile-quantile plots. Variants with an imputation quality metric of < 0.3 (provided a higher threshold for imputation quality was not already implemented), a minor allele count (MAC) of < 20 using the effective sample size or the number of cases and adjusted for imputation quality where applicable, were set to missing. Variants were included for meta-analysis if they were present in at least 13 studies (those with European ancestry and at least 7 million markers passing all quality control filters).

Staged GWAS meta-analysis

In Stage 1 of the analysis, we used Metal[56,57] version 2011-03-25 to perform a fixed-effects meta-analysis of genome-wide data from 22 studies and four additional COPD cohorts genotyped on an Illumina HumanExome v1.2 platform with custom content; this content included a set of COPD candidate genes and regions identified from prior COPD GWAS efforts[13]. We adjusted for inflation using genomic control correction in each study. We included study populations with subjects of non-European ancestry in the overall analysis, and additionally examined results limited to study populations of European ancestry. To identify variants to test for association in Stage 2 in the UK BiLEVE study, we selected top results ($P < 5x10^{-6}$) from the Stage 1 meta-analysis. We selected one lead variant from the chromosome 15q25, *FAM13*, and *HHIP* regions, as all of these have been described in multiple COPD GWASs[13,16,17,21]. For the remainder of the regions, we performed linkage disequilibrium pruning using the PLINK1.9 --clump procedure with an r^2 of 0.5, additionally examining these SNPs for the number of cohorts with passing quality control at each variant and including SNPs in strong LD (i.e., part of the same clump) with

a lower degree of missingness. To identify independent results, we used GCTA-COJO[58,59] on the Stage 1 meta-analysis for variants with $P < 5x10^{-6}$ using the default distance of 10Mb. We used the COPDGene non-Hispanic whites (as the largest representative population) as the reference population for these analyses. An overall meta-analysis across the Stage 1 and Stage 2 (UK BiLEVE) cohorts was performed and variants with $P < 5x10^{-8}$ were considered genome-wide significant (**Figure 5-1**).

Lung eQTL analysis

Lung expression quantitative trait loci (eQTL) were calculated from 1,111 human subjects who underwent lung surgery at three academic sites, Laval University, University of British Columbia (UBC), and University of Groningen, henceforth referred to as Laval, UBC, and Groningen, respectively. This lung eQTL dataset has been described previously[22,60]. Briefly, 66.7% to 91.2% of the individuals in this study were current or former smokers and 24.2% to 35.3% had moderate to severe COPD (GOLD spirometry grade 2 to 4). Whole-genome gene expression profiling in the lung was performed on a custom Affymetrix array (GPL10379). Microarray pre-processing and quality controls were described previously[22,61,62]. Probe sequences were mapped to the human genome (hg19) using Bowtie[63] and probes not mapping to a coding region or having a common SNP (MAF ≥ 5%) in their sequence were removed. Expression data were adjusted for age, sex, and smoking status using residuals obtained with the robust fitting of linear models function (rlm) in the R statistical package MASS. Residual values deviating from the median by more than three standard deviations were filtered as outliers. Genotyping was carried on the Illumina Human 1M-Duo BeadChip array.

Twenty-one out of the 22 SNPs (in main manuscript **Table 5-1**) were genotyped or imputed in the three cohorts, i.e. Laval, UBC, and Groningen. One of the SNPs, rs7186831, was not well-imputed; a proxy, rs11865296 in modest linkage disequilibrium ($r^2 = 0.54$, 1000 genomes phase 3, EUR) was used instead. These variants were tested for association with adjusted expression traits (43,465 probe sets) in the lung. SNPs within 1 Mb up and downstream of the transcription probe set were considered as local-eQTL. Distant-acting eQTLs were further than 1 Mb away or on a different chromosome. Association tests were carried with PLINK1.9[64,65] in each cohort and then meta-analyzed using Fisher's method.

All local eQTL with nominal P value < 0.05 in the meta-analysis were considered. To provide an additional overall estimate of eQTL significance, we considered a Bonferroni correction threshold ([0.05/(22 SNPs x 43,465 probe sets) = P value < 5.2×10^{-8}]). Statistical analyses were performed in R3.2.3[66].

Co-localization Analysis

Co-localization of statistical signals between COPD genetic association and eQTL were examined using the coloc R package[23]. We used phenotypic summary statistics from individuals of European ancestry with genome-wide association data and all eQTL results and examined 500kb flanks around the top 22 genome-wide significant associations found in the overall meta-analysis (**Table 5-1**).

Sensitivity Analysis

To estimate the effect of using pre- instead of post-bronchodilator lung function on our results, we examined the top set of genome-wide significant results in our largest cohorts with both pre- and post-bronchodilator data and densely imputed genotypes (COPDGene NHW and AA, ECLIPSE, NETT-NAS, and Norway / GenKOLS). Since subjects from these cohorts (except for COPDGene) were included based on post-bronchodilator values, including all subjects with COPD based on post-bronchodilator spirometry would lead to larger sample sizes and make comparison of P-values more difficult. Thus, we chose a random sample of post-bronchodilator cases and controls that matched the number of pre-bronchodilator cases and controls. We performed logistic regression using these equal sized set of pre- and post-bronchodilator cohorts, and meta-analyzed the results.

Asthma overlap analysis

We assessed the overlap between our results and known asthma susceptibility loci. We downloaded information on genome-wide significant ($P < 5 \times 10^{-8}$) associations with asthma and asthma-related traits including asthma and hay fever, asthma (childhood onset), asthma (corticosteroid response), bronchodilator response in asthma, pulmonary function decline, and severe asthma in the NHGRI-EBI GWAS Catalog[42]. Additionally, we

examined top associated variants (which were not genome-wide significant) in the susceptibility to the asthma-COPD overlap syndrome[67]. In all, we assessed the association statistics of 49 unique asthma-associated trait loci across 26 genomic regions in our Stage 1 meta-analysis results. We also examined the asthma association statistics of our top COPD loci from overall meta-analysis using publically available asthma GWAS data from the GABRIEL Consortium[43]. For COPD loci not present in the GABRIEL Consortium asthma GWAS data, we attempted to examine proxy SNPs in LD ($r^2 > 0.5$, 1000 genomes phase 1 CEU) with our top COPD loci.

To examine the genetic correlation[47] of COPD and asthma over the entire genome, we performed LD score regression[46] using summary statistics from publically available asthma GWAS data from the GABRIEL Consortium[43]. For all comparisons using LD score regression, we filtered to HapMap3 variants, limited to European-ancestry subjects with genome-wide data, and filtered on missingness using default parameters in munge_sumstats.py. For the GABRIEL data, we required a variant to be present in at least 35 of the studies.

Smoking behavior overlap analysis

We downloaded information on genome-wide significant ($P < 5x10^{-8}$) associations with the traits "nicotine dependence" and "smoking behaviour" in the NHGRI-EBI GWAS Catalog[42]. We assessed these top smoking-associated SNPs in our Stage 1 meta-analysis results. We also assessed overlap of smoking and COPD in the publically available summary statistics from the 2010 Tobacco and Genetics Consortium GWAS[45]. We evaluated our top COPD loci associations from overall meta-analysis with both cigarettes per day and ever-smoking traits. For COPD risk SNPs not directly analyzed in the Tobacco and Genetics Consortium GWAS, we attempted to examine proxy SNPs in LD ($r^2 > 0.5$, 1000 genomes phase1 CEU) with our top COPD loci.

To examine the genetic correlation[47] of COPD and smoking behaviours (cigarettes per day and ever-smoking status) over the entire genome, we performed LD score regression[46] using summary statistics from our current COPD study as noted above and

publically available summary statistics from the 2010 Tobacco and Genetics Consortium GWAS[45].

Fine mapping analysis

We attempted to determine, at each locus, whether we could identify a potentially causal variant. We performed these analyses using European ancestry subjects alone, and in all subjects with genome-wide data, and excluded variants that were not present in at least 80% of the full sample. We assumed a single causal variant at each locus, examined a +/- 250kb region around the top variant, and calculated approximate Bayes factors using the method of Wakefield[68] to determine the 95% credible set. While specific trans-ethnic mapping approaches[69,70] can significantly assist in identifying causal loci, we found that the number of non-European samples in our study were likely insufficient to leverage these methods.

Functional enrichment analysis

To identify enriched cell types for our COPD associations, we applied LD score regression to GenoSkyline[71] lung tissue annotations (the default LD score regression annotations collapse lung into the cardiovascular tissue type), as well as cell-type specific annotations from LD score regression[46]. We also performed analysis using only the 22 genome-wide significant loci and tested for enrichment of imputed chromatin marks from ROADMAP using HaploReg 4.1[29]. Further, we applied a more sophisticated analysis adjusting for local linkage disequilibrium patterns, GoShifter[72]. Finally, we examined overlap with gene expression datasets using SNPsea[73].

Additional pulmonary fibrosis and COPD overlap analysis

To further examine overlapping loci for COPD and pulmonary fibrosis, we combined summary statistics from our study and the pulmonary fibrosis GWAS by Fingerlin et al.[8,9] using gwas-pw[37].

Causal gene analysis

For the genome-wide significant loci from the overall meta-analysis, we explored potential causative genes at each association locus using the PrixFixe method[74], assuming co-function of all significant loci. As required by the PrixFixe method, we assured our genome-wide significant loci were present in dbSNP v137[75] and were represented HapMap[76] phase III data; for loci not meeting these requirements, proxy SNPs from HapMap phase III were selected based on strongest LD (r^2) with index SNP (see **Supplementary Table 5-S11** for details of the proxy variant used at each genome-wide significant locus).

COPD comorbidity overlap analysis

We assessed the overlap between our results and two common COPD comorbidities, coronary artery disease and osteoporosis (through bone mineral density traits). We downloaded information on genome-wide significant ($P < 5x10^{-8}$) associations with these comorbidities as reported in the NHGRI-EBI GWAS Catalog[42]. We assessed the association statistics of these comorbid trait loci in our Stage 1 meta-analysis results.

Quantitative imaging overlap analysis

To explore the relationship between our top COPD-associated variants and imaging features of emphysema and airway thickness, we queried data from a GWAS of COPD quantitative imaging features[77]. For each genome-wide significant COPD susceptibility locus in our overall meta-analysis, we assessed the corresponding quantitative imaging GWAS effect size, effect direction, and P value for association with the following quantitative imaging traits: %LAA-950 (percentage low attenuation area, using a threshold of -950 Hounsfield units); Perc15 (value of Hounsfield units at the 15th percentile of the density histogram); Pi10 (airway wall area: the value for a hypothetical airway of 10 mm internal perimeter obtained by plotting a regression line of the square root of the airway wall area versus the airway internal perimeter); and WAP (percentage of the wall area compared to the total bronchial area).

Gene set enrichment analysis

As an attempt to minimize false positives in our gene set enrichment analysis, we divided the Stage 1 GWAS cohorts with full genome-wide data into two sets of roughly equal size. We then used i-GSEA4GWAS (http://gsea4gwas.psych.ac.cn)[78] for each of the two GWAS data sets to assess enrichment of COPD GWAS loci in BioCarta (http://cgap.nci.nih.gov/Pathways/BioCarta_Pathways) and KEGG[79] pathways as well as gene ontology (GO) terms[80,81]. We first evaluated GO terms and pathways with a false-discovery rate (FDR) less than 5% in both analysis sets and then used a more stringent threshold of FDR < 1% to evaluate overlap of GO term and pathway enrichment in our two analysis sets.

References

54. Castaldi, P.J. *et al.* The association of genome-wide significant spirometric loci with chronic obstructive pulmonary disease susceptibility. *Am J Respir Cell Mol Biol* 45, 1147-53 (2011).

55. Winkler, T.W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* 9, 1192-212 (2014).

56. Abecasis, G., Li, Y. & Willer, C. METAL MetaAnalysis Helper. Version release 2011-03-25. URL: http://genome.sph.umich.edu/wiki/METAL_Program. (2011).

57. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26, 2190-1 (2010).

58. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 88, 76-82 (2011).

59. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* 44, 369-75, S1-3 (2012).

60. Lamontagne, M. *et al.* Refining susceptibility loci of chronic obstructive pulmonary disease with lung eqtls. *PLoS One* 8, e70220 (2013).

61. Bosse, Y. *et al.* Molecular signature of smoking in human lung tissues. *Cancer Res* 72, 3753-63 (2012).

62. Lamontagne, M. *et al.* Genetic regulation of gene expression in the lung identifies CST3 and CD22 as potential causal genes for airflow obstruction. *Thorax* 69, 997-1004 (2014).

63. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25 (2009).

64. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81, 559-75 (2007).

65. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7 (2015).

66. R Core Team (2016). R: A language and environment for statistical computing. (R Foundation for Statistical Computing. URL: http://www.R-project.org/, Vienna, Austria).

67. Hardin, M. *et al.* The clinical and genetic features of COPD-asthma overlap syndrome. *Eur Respir J* 44, 341-50 (2014).

68. Wakefield, J. Bayes factors for genome-wide association studies: comparison with P-values. *Genet Epidemiol* 33, 79-86 (2009).

69. Morris, A.P. Transethnic meta-analysis of genomewide association studies. *Genet Epidemiol* 35, 809-22 (2011).

70. Kichaev, G. & Pasaniuc, B. Leveraging Functional-Annotation Data in Trans-ethnic Fine-Mapping Studies. *Am J Hum Genet* 97, 260-71 (2015).

71. Lu, Q., Powles, R.L., Wang, Q., He, B.J. & Zhao, H. Integrative Tissue-Specific Functional Annotations in the Human Genome Provide Novel Insights on Many Complex Traits and Improve Signal Prioritization in Genome Wide Association Studies. *PLoS Genet* 12, e1005947 (2016).

72. Trynka, G. *et al.* Disentangling the Effects of Colocalizing Genomic Annotations to Functionally Prioritize Non-coding Variants within Complex-Trait Loci. *Am J Hum Genet* 97, 139-52 (2015).

73. Slowikowski, K., Hu, X. & Raychaudhuri, S. SNPsea: an algorithm to identify cell types, tissues and pathways affected by risk loci. *Bioinformatics* 30, 2496-7 (2014).

74. Tasan, M. *et al.* Selecting causal genes from genome-wide association studies via functionally coherent subnetworks. *Nat Methods* 12, 154-9 (2015).

75. Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda, MD: National Center for Biotechnology Information, National Library of Medicine. (dbSNP Build ID: 137). URL: http://www.ncbi.nlm.nih.gov/SNP/.

76. International HapMap, C. The International HapMap Project. *Nature* 426, 789-96 (2003).

77. Cho, M.H. *et al.* A Genome-Wide Association Study of Emphysema and Airway Quantitative Imaging Phenotypes. *Am J Respir Crit Care Med* 192, 559-69 (2015).

78. Zhang, K., Cui, S., Chang, S., Zhang, L. & Wang, J. i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res* 38, W90-5 (2010).

79. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28, 27-30 (2000).

80. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25, 25-9 (2000).

81. Gene Ontology, C. Gene Ontology Consortium: going forward. *Nucleic Acids Res* 43, D1049-56 (2015).

Supplementary Materials

See the attached file.

CHAPITRE 6

Comprehensive genetic testing for alpha-1 antitrypsin deficiency in COPD patients

Comprehensive genetic testing for alpha-1 antitrypsin deficiency in COPD patients

Maxime Lamontagne[1], MSc; Nathalie Gaudreault[1], BSc; François Maltais[1], MD; Wan Tan[2], MD; Jean Bourbeau[3], MD; Yohan Bossé[1,4], PhD; for the Canadian Respiratory Research Network, for the CanCOLD study group

1) Institut universitaire de cardiologie et de pneumologie de Québec, Québec, Québec, Canada.

2) University of British-Columbia, Vancouver, British-Columbia, Canada.

3) McGill University, Montréal, Québec, Canada.

4) Department of Molecular Medicine, Laval University, Québec, Canada.

Address correspondence to:

Yohan Bossé, Ph.D.

Institut universitaire de cardiologie et de pneumologie de Québec

Pavillon Marguerite-d'Youville, Y2106

2725, chemin Sainte-Foy

Québec (Québec)

Canada, G1V 4G5

email: yohan.bosse@criucpq.ulaval.ca

Summary conflict of interest statements: No conflicts exist for the authors

Funding information

This work was supported by the *Fondation de l'Institut universitaire de cardiologie et de pneumologie de Québec*, the Respiratory Health Network of the *Fonds de recherche Québec – Santé* (FRQS), the Canadian Institutes of Health Research (MOP - 123369), and the Canadian Respiratory Research Network (CRRN). The CRRN is supported by grants from the Canadian Institutes of Health Research (CIHR) - Institute of Circulatory and Respiratory Health; Canadian Lung Association (CLA)/Canadian Thoracic Society (CTS); British Columbia Lung Association; and Industry Partners Boehringer-Ingelheim Canada Ltd, AstraZeneca Canada Inc., and Novartis Canada Ltd. Funding for training of graduate students and new investigators within the network was supported by the above funding Sponsors and by GlaxoSmithKline Inc. M Lamontagne is the recipient of a doctoral studentship from the FRQS. F Maltais holds a GSK/CIHR Research Chair on COPD at Université Laval. Y Bossé holds a Canada Research Chair in Genomics of Heart and Lung Diseases.

Abbreviations

AAT = alpha-1 antitrypsin; AATD = alpha-1 antitrypsin deficiency; CanCOLD = The Canadian Cohort of Obstructive Lung Disease; COPD = chronic obstructive respiratory disease; Pi = protease inhibitor; *SERPINA1* = serpin peptidase inhibitor clade A member 1

Résumé

Introduction : Le déficit en alpha-1 antitrypsine (DAAT) est une maladie héréditaire causée par des mutations dans le gène *SERPINA1*. Les individus atteints ont un risque augmenté de souffrir d'emphysème pulmonaire, particulièrement ceux ayant des antécédents de tabagisme. L'algorithme présentement employé en clinique pour identifier le DAAT comprend plusieurs étapes et utilise des techniques désuètes.

Objectif : Ce projet vise à vérifier la faisabilité et l'utilité du séquençage d'ADN afin d'optimiser la détection du DAAT dans une cohorte de patients canadiens avec MPOC.

Méthodes : Le séquençage des régions codantes du gène *SERPINA1* fut effectué dans une cohorte de 400 patients MPOC provenant de l'étude *Canadian Cohort of Obstructive Lung Disease* (*CanCOLD*).

Résultats : Au total, 19 variations génétiques, dont 15 mutations faux sens, ont été identifiées dans le gène *SERPINA1*. Parmi les mutations observées, trois génèrent des niveaux normaux d'alpha-1 antitrypsine (Pi M1 à M4), cinq sont connues pour causer un déficit léger à sévère (Pi Z, P_{Lowell}, F, I et S), et sept ont des effets inconnus. Trois individus ont été classifiés DAAT incluant deux homozygotes ZZ et un hétérozygote composé SZ. Soixante-et-onze patients (18%) étaient porteurs d'au moins un allèle déficient. Le séquençage d'ADN s'effectue à l'intérieur d'un délai de 48 heures, est peu coûteux et permet de déterminer de façon définitive la nature de l'anomalie génétique sous-jacente.

Conclusion : Le séquençage d'ADN offre plusieurs avantages comparativement à l'algorithme à plusieurs étapes actuellement employé en clinique. L'utilisation de cette technique permettrait d'accélérer le diagnostic et diminuer les coûts et ainsi faciliter la prise en charge et le traitement des patients DAAT en fonction des recommandations des Sociétés de thoracologie sur les tests ciblés et le traitement substitutif.

Abstract

Background: Alpha-1 antitrypsin deficiency (AATD) is one of the most common hereditary disorders. AATD is caused by genetic mutations in the *SERPINA1* gene and is associated with increased risk of lung and liver diseases. Only a minority of AATD patients is identified and the current multi-step diagnosis algorithms are not helping to enhance recognition of this condition.

Objective: The goal of this study is to evaluate the feasibility and effectiveness of DNA sequencing of the *SERPINA1* gene as a single test to diagnose AATD.

Methods: DNA sequencing of the coding regions of *SERPINA1* was performed as a single diagnostic test to detect AATD. The frequencies of deficient alleles were evaluated in 400 Canadian patients with chronic obstructive respiratory disease (COPD).

Results: Nineteen genetic variants were identified including 15 missense mutations. These include three mutations known to generate normal protease inhibitor (Pi) M1 to M4 alleles; five causing deficient Pi alleles Z, S, P_{Lowell}, F, and I; and seven with unknown significance. Overall, at least 9.4% of alleles in this Canadian COPD cohort were deficient. Three patients were identified as AATD including two homozygotes ZZ and one compound heterozygote SZ. Seventy-one out of 400 patients (18%) were carriers of at least one deficient allele. DNA sequencing can be easily implemented, with a turnaround of less than 48 hours, and a price that is lower than the multi-step algorithms currently used to diagnose AATD.

Conclusions: DNA sequencing of *SERPINA1* is effective as a single test strategy to detect AATD. Simplifying and improving diagnostic strategy will ultimately expedite the clinical decision-making process recommended by many national and international lung societies about targeted testing and augmentation therapy.

Introduction

Alpha-1 antitrypsin deficiency (AATD) is an inherited disorder associated with accelerated rate of lung function decline and early onset emphysema. AATD is caused by genetic mutations in the *SERPINA1* gene located on chromosome 14, which encodes an antiprotease called alpha-1 antitrypsin (AAT). The prevalence of AATD is estimated at one in 3,000 to 5,500 in the North American population[1]. Previous studies suggested that severe AATD accounts for 1% to 5% of COPD cases[2-4], but the exact number for this specific population remains to be described. Clinicians frequently under recognize AATD and most cases are not diagnosed. The widespread under recognition of AATD may be related to the low awareness of this condition among physicians, incomplete penetrance of the genetic defect, clinical manifestations that are not exclusive to this condition, but also to the complexity of diagnosis[4]. This results in an unacceptable delay before the diagnosis is reached[5]. In Canada, a typical patient will experience a diagnostic delay of 7 years after the onset of respiratory symptoms[6], which is not different from elsewhere[5].

Diagnostic methods for AATD vary by countries and regions, but usually follow a multi-step testing algorithm[7-9]. First, measuring the serum or plasma levels of AAT is typically used as an initial test in individuals with suspected deficiency. However, this strategy does not provide information about the genetic variants involved. In addition, acute inflammatory/infectious conditions may lead to falsely elevated AAT levels[10]. Second, targeted genotyping is considered with or independently of serum AAT measurements. Commercially available genotyping tests developed so far are testing only the most common mutations (S and Z) and not all known genetic variants in *SERPINA1*. This is a major limitation considering that over 250 gene variants have been reported[11] [Exome Aggregation Consortium (ExAC), Cambridge, MA (URL: http://exac.broadinstitute.org) accessed November, 2015]. Third, presumptive genotypes of AAT can also be obtained by protease inhibitor (Pi) phenotyping. Pi typing is determined by isoelectric focusing of serum AAT and reflects the genotype at the *SERPINA1* locus[7]. There are four main variants of AAT classified by their speed of movement during gel electrophoresis and labeled as fast (F), medium (M), slow (S), and very slow (Z)[12]. However, Pi typing only provides presumptive genotypes, has limited resolution (i.e. characterize a limited number of protein phenotypes), cannot detect heterozygote carriers of null variants, requires correlation with

the individual's serum AAT level, and the outcome is sometime difficult to interpret. Finally, the gold standard to detect the deficiency is direct DNA sequencing. Historically, full-gene sequencing is considered laborious, expensive and not available in all centers[13]. In the current genomic era, these arguments are no longer valid and the time has come to develop a comprehensive genetic test for AATD and replace the multi-step testing algorithms into a single test that is inexpensive, fast, and accurate.

The goals of this study were two folds. First evaluate the feasibility and effectiveness of DNA sequencing of the *SERPINA1* gene as a single test to diagnose AATD in order to expedite the clinical decision-making process recommended by many organizations about targeted testing and augmentation therapy[4,13]. Second, test the frequencies of AATD alleles in a Canadian COPD population.

Materials and Methods

CanCOLD study

The Canadian Cohort of Obstructive Lung Disease (CanCOLD) is a prospective multi-center cohort study in Canada specifically designed to increase our understanding of COPD and to improve the management of affected patients[14]. Importantly, CanCOLD represents a random sampling of the Canadian population. So far, 1,552 individuals over the age of 40 living in 10 cities across Canada were enrolled. All research participants were thoroughly phenotyped for COPD and related variables including sociodemographic, clinical, pulmonary function and CT scans of the chest. Three visits are scheduled at baseline, 18 months and 3 years. Only data from the initial visit is available for the current study. A subset of 400 CanCOLD participants with COPD was selected for DNA sequencing. The patients with the most severe COPD (GOLD 2+) and milder COPD with earlier age of onset were prioritized. **Table 6-1** shows the clinical characteristics of these COPD patients.

DNA sequencing of *SERPINA1*

DNA was extracted from 200 µL of frozen whole blood using QIAamp® DNA Blood Mini kit (QIAGEN, Mississauga, ON, Canada). The DNA quality and concentration was assessed by the UV absorbance ratio 260 nm/280 nm and UV absorbance 260 nm, respectively. Every sample was diluted to a final concentration of 50 ng/µL. The DNA sequences of the coding regions (i.e. exons 2 to 5) of the *SERPINA1* gene were obtained by Sanger sequencing in 400 COPD patients. Primer sequences to evaluate the selected regions of *SERPINA1* are provided in **Table 6-2**. PCR was performed in a final volume of 25 µl containing 100 ng of genomic DNA, 1 U of HotStar Taq DNA polymerase (Qiagen), PCR buffer 1X, Q-Solution 1X, 160 µM of each dNTP and 0.2 µM of each primer. The PCR reaction was carried out on the GeneAmp® PCR system 9700 (Applied Biosystems) with the following cycling conditions: 15 minutes at 95ºC, 35 PCR amplification cycles (15 seconds at 94ºC, 30 seconds at 60ºC, and 60 seconds at 72ºC), and 7 minutes at 72ºC. The sequencing reaction was then performed using standard procedures and the product was run on the ABI 3730xl DNA Analyzer (Applied Biosystems). Sequencing files were assembled and analyzed using the EMBL-EBI Clustal Omega Multiple Alignment Tool (http://www.ebi.ac.uk/Tools/msa/clustalo).

Results

DNA sequencing of *SERPINA1*

One important goal of this project was to evaluate DNA sequencing as a single test for the detection of AATD. The idea was to provide a diagnostic test at the DNA-based resolution, instead of relying on fluctuating levels of AAT protein in serum and the migrating velocity of the protein on gel. **Table 6-3** describes time and cost (in Canadian dollars) to perform the test and the different components of the test including DNA extraction, quality control, PCR amplification, DNA sequencing, and data analysis. Please note that these metrics are estimated for a general genetic laboratory with available equipment including a thermocycler and a DNA sequencing system. As an example, time is estimated based on one laboratory technician processing one batch of 24 samples, which is

a convenient number to work with considering the 96-well DNA plate format. Overall, the process starting with blood takes approximately 13 hours and can be easily implemented in two working days. It should be emphasized that the amplification (PCR) and sequencing runs are automatic and require no hands-on time. Accordingly, the process takes about 8 hours of hands-on time for an experienced laboratory technician. Reagent costs for 24 samples is estimated at $680, or approximately 28$ per patient. Adding cost associated with staff member (13 hours x $30/hour = $390) will raise the cost to roughly 45$ per patient. Overall, this genetic test can diagnose AATD at the DNA-based resolution for a cost price of approximately 50$ CDN and with a turnaround of less than 48 hours. The volume of samples processed at the same time will influence cost and time. These metrics are likely to rise for smaller laboratory and decrease for higher throughput facilities.

Frequencies of genetic variants and AATD alleles in a Canadian COPD population

DNA sequencing of *SERPINA1* was performed in 400 COPD patients from CanCOLD. Briefly, 19 genetic variants were identified (**Figure 6-1**). **Table 6-4** shows the variants based on standardized genetic, nucleotide, protein, and Pi typing nomenclatures as well as allele frequencies among the 400 COPD patients. The allele frequencies of normal alleles in this population was 17.8% for M1 (Ala^{213}), 45.5% for M1 (Val^{213}), 13.9% for M2, 11.1% for M3, and 2.4% for M4. Accordingly, 90.6% of alleles were normal.

A number of genetic variants responsible for AAT phenotypes defined by isoelectric focusing of serum were identified. The most common was rs17580 causing the S phenotype with an allele frequency of 6.4%. Rs28929474 causing the Z phenotype was found at 2.4%. Rs28931570 and rs121912714 causing the I and P_{Lowell} phenotypes, respectively, were identified with an allele frequency of 0.25%. Finally, rs28929470 responsible for the F phenotype was found with an allele frequency of 0.13%. The S, Z, I, P, and F are known deficient alleles. Taking together, 75 deficient alleles were identified and indicated that at least 9.4% of alleles (75 out of 800) in this Canadian COPD cohort are deficient.

Additional missense genetic variants were identified, for which the AAT effect and clinical relevance are not known. This includes four variants in exon 2: rs745463238 (Ser14Phe), rs111850950 (Ala60Thr), rs758820515 (Pro88Leu), and rs112030253 (Gly148Trp). One more missense in exon 3 (rs759736224, Pro255Thr) and two in exon 4 (rs141620200, Aal284Ser; rs201788603, Ser330Phe) were identified. The allele frequencies for all these missense variants varied from 0.13% to 0.25% (one or two heterozygote patients). Collectively, these variants account for 1.1% of all alleles. Interestingly, the Ser14Phe mutation was identified in one individual with PiSZ genotype, indicating that this mutation occurs on the background of the S or Z alleles.

One newly identified variant was found and labelled c.738G>A based on standard nomenclature[15]. This variant is located in exon 3 and causes a G to A substitution at position 3 of codon 222 (AAG to AAA). This new variant is synonymous, i.e. does not change the lysine at position 222 of the protein, and is thus unlikely to have a functional impact. This new variant occurred on the M1 (Val^{213}) background.

Frequencies of AATD and carriers of deficient alleles in a Canadian COPD population

Two patients had genotype AA for SNP rs28929474, meaning that these individuals carry two copies of the mutant Z allele and are thus confirmed cases of AATD. One patient was compound heterozygote SZ carrying allele A for SNP rs28929474 and allele T for SNP rs17580. Fourteen additional patients were heterozygotes for the Z allele. One SS individual was found and 49 additional patients carried the S allele. One patient was a carrier of the F allele with a T for SNP rs28929470. Two heterozygotes carriers of deficient alleles I (rs28931570-T) and P (rs121912714-T) respectively were also found. **Figure 6-2** illustrates the number of individuals in any combinations of normal and deficient alleles. Overall, 71 out of 400 patients (18%) were carriers of at least one deficiency allele. This includes three severely deficient (2 PiZZ and 1 PiSZ) and 68 mild to moderate deficient (2 PiMP, 2 PiMI, 1 PiMF, 1 PiSS, 14 PiMZ, and 48 PiMS).

Discussion

Alpha-1 antitrypsin deficiency (AATD) is one of the most common hereditary disorders in the world, with more than one million cases with PiZZ and PiSZ genotypes[1]. International and national lung societies and most particularly patient groups have had a tremendous contribution during the last 20 years to increase awareness about AATD. Unfortunately, practice gaps persist despite clear recommendations about appropriate patients suitable for targeted testing for AATD[4,13]. For example, the ATS/ERS statement recommended diagnostic testing for adults with persistent obstructive defects including emphysema, COPD, and asthma in which airflow obstruction is incompletely reversible as well as individuals with unexplained liver disease. Predispositional testing is also recommended for individuals with a family member with AATD. Why is it that only a limited number of clinicians are following these recommendations? One important bottleneck at many locations is the lack of a fast and accurate diagnostic test. In addition, most of the well-established laboratories for AATD are using outdated techniques to detect AATD. In the current era of genomic medicine, the diagnosis of AATD must go beyond the PiZ and S deficient alleles. A diagnosis at the DNA-based resolution is important to distinguish genetic and nongenetic (e.g. hepatic dysfunction) AATD as well as to determine a proper prognosis as some genetic variants are associated with pulmonary symptoms only while others cause pulmonary and hepatic symptoms. In addition, a definite genetic diagnosis would raise the level of awareness of the range of deficiency and stimulate further studies into their clinical relevance. In this study, we have used DNA sequencing to provide a complete assessment of the gene causing AATD, namely *SERPINA1*. DNA sequencing is providing diagnosis at the DNA-based level, which is the ultimate level of information needed for hereditary disorders. DNA sequencing can detect conventional Pi deficient alleles (e.g. Z and S), but also rare and novel genetic variants. Most importantly, DNA sequencing can be implemented in any laboratory with a standard DNA sequencing system, has a turnaround of less than 48 hours, and is cheaper than the multi-step testing algorithms used in most countries.

In this Canadian COPD cohort, 0.75% of patients were AATD, i.e. PiZZ and SZ, which is lower than 1.9% of PiZZ found in 965 severe COPD patients[2]. Random sampling

in this study[14] and patients with a range of COPD severity may explained this difference. Despite this lower than expected ZZ homozygosity, deficient alleles together including Z, S, I, F and P_{Lowell}, account for nearly 10% of all alleles tested (75 out of 800). Overall, these alleles affect 18% of patients that have thus a genetic predisposition to develop COPD. Seven additional missense mutations were also identified, each in one or two patients. Unfortunately, the frequency of these mutations is too low to draw conclusions about their clinical significance and generate uncertainty for the diagnosis of nine additional patients (2.25%). One patient was a carrier of the F allele, which confers a dysfunctional protein with normal AAT blood levels[16]. This case further highlights the intrinsic limitation of relying on serum measurement of AAT and commercially available genotyping tests to establish the diagnosis of AATD. Overall, these results demonstrated the effectiveness of a diagnostic test performed at the DNA-based resolution. Studying the effects of known and newly identified genetic variants with serum levels of AAT and the clinical phenotypes of patients in this COPD cohort will be part of a future study.

Registries of patients with AATD have been very valuable to carry out clinical trials and to understand the natural history of the disease[17]. For most national registries, the diagnostic of AATD is restricted to serum AAT levels and/or the most common mutations (PiS and Z). A comprehensive diagnostic test at the DNA-based level is needed to improve our ability to combine results across registries and understand the clinical relevance of rare mutations. This study highlights the importance of testing less frequent deficient alleles. Together, these rare alleles represent a significant number of patients with the potential of mislabeling their AATD status with the current testing procedure. This study demonstrated that DNA sequencing is fast, cheap, and accurate, suggesting that the time has come to replace the historical multi-step diagnostic algorithms used by most countries [7-9] by a single-step DNA sequencing test. Consistency in diagnosis at the DNA-based resolution has the potential to accelerate outcomes of existing registries, expedite the clinical decision-making process recommended by many national and international lung societies for targeted testing and augmentation therapy[4,13], and ultimately, improve quality of care for AATD patients.

Author contributions

Y.B. had full access to the data in the study and takes responsibility for the integrity of the data and article. M.L. contributed to DNA sequencing analysis. N.G. performed the DNA sequencing. F.M., W.T., and J.B. contributed to the study design, performed clinical recruitment and phenotyping of patients. Y.B. conceived the study, contributed to DNA sequencing analysis and interpretation, and wrote the manuscript. M.L., N.G., F.M., W.T., and J.B. edited the manuscript for intellectual content and approved the final manuscript.

Financial/nonfinancial disclosures

None

Role of the sponsors

The funders had no role in the study design, data collection and analysis, or preparation of the manuscript.

References

1. de Serres FJ. Worldwide racial and ethnic distribution of alpha1-antitrypsin deficiency: summary of an analysis of published genetic epidemiologic surveys. *Chest.* 2002;122(5):1818-1829.

2. Lieberman J, Winter B, Sastre A. Alpha 1-antitrypsin Pi-types in 965 COPD patients. *Chest.* 1986;89(3):370-373.

3. Abboud RT, Ford GT, Chapman KR. Alpha1-antitrypsin deficiency: a position statement of the Canadian Thoracic Society. *Can Respir J.* 2001;8(2):81-88.

4. American Thoracic Society/European Respiratory Society statement: standards for the diagnosis and management of individuals with alpha-1 antitrypsin deficiency. *Am J Respir Crit Care Med.* 2003;168(7):818-900.

5. Stoller JK, Sandhaus RA, Turino G, Dickson R, Rodgers K, Strange C. Delay in diagnosis of alpha1-antitrypsin deficiency: a continuing problem. *Chest.* 2005;128(4):1989-1994.

6. Bradi AC, Audisho N, Casey DK, Chapman KR. Alpha-1 antitrypsin deficiency in Canada: regional disparities in diagnosis and management. *COPD.* 2015;12 Suppl 1:15-21.

7. Snyder MR, Katzmann JA, Butz ML, et al. Diagnosis of alpha-1-antitrypsin deficiency: An algorithm of quantification, genotyping, and phenotyping. *Clin Chem.* 2006;52(12):2236-2242.

8. Ferrarotti I, Scabini R, Campo I, et al. Laboratory diagnosis of alpha1-antitrypsin deficiency. *Transl Res.* 2007;150(5):267-274.

9. Graham RP, Dina MA, Howe SC, et al. SERPINA1 Full-Gene Sequencing Identifies Rare Mutations Not Detected in Targeted Mutation Analysis. *J Mol Diagn.* 2015;17(6):689-694.

10. Ferrarotti I, Thun GA, Zorzetto M, et al. Serum levels and genotype distribution of alpha1-antitrypsin in the general population. *Thorax.* 2012;67(8):669-674.

11. DeMeo DL, Silverman EK. Alpha1-antitrypsin deficiency. 2: genetic aspects of alpha(1)-antitrypsin deficiency: phenotypes and genetic modifiers of emphysema risk. *Thorax.* 2004;59(3):259-264.

12. Fagerhol MK, Laurell CB. The Pi system-inherited variants of serum alpha 1-antitrypsin. *Prog Med Genet.* 1970;7:96-111.

13. Marciniuk DD, Hernandez P, Balter M, et al. Alpha-1 antitrypsin deficiency targeted testing and augmentation therapy: a Canadian Thoracic Society clinical practice guideline. *Can Respir J.* 2012;19(2):109-116.

14. Bourbeau J, Tan WC, Benedetti A, et al. Canadian Cohort Obstructive Lung Disease (CanCOLD): Fulfilling the need for longitudinal observational studies in COPD. *COPD.* 2014;11(2):125-132.

15. den Dunnen JT, Antonarakis SE. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat.* 2000;15(1):7-12.

16. Ringenbach MR, Banta E, Snyder MR, Craig TJ, Ishmael FT. A challenging diagnosis of alpha-1-antitrypsin deficiency: identification of a patient with a novel F/Null phenotype. *Allergy Asthma Clin Immunol.* 2011;7(1):18.

17. Stockley RA. Antitrypsin Deficiency Assessment and Programme for Treatment (ADAPT): The United Kingdom Registry. *COPD.* 2015;12 Suppl 1:63-68.

Tables

Table 6-1 : Clinical characteristics of COPD patients.

Characteristics	
Number of patients	400
Age (years; mean ± SD)	62.0 ± 7.5
Sex – no. of patients (%)	
Male	237 (59)
Female	163 (41)
COPD GOLD Stage – no. of patients (%)	
1	172 (43)
2+	228 (57)

Table 6-2 : Primers to sequence the SERPINA1 gene.

	Forward	Reverse	Amplicon Size (bp)
Exon 2	5'- CAAGTACTTGGCACAGGCTGG -3'	5'- GGCCCATAATGCATTGCCAAGG -3'	873
Exon 3	5'- GAAGCTGAGCCTCGAGGGATG -3'	5'- AGCAGTGACCCAGGGATGTG -3'	533
Exon 4	5'- GAGGACACAGGAGTAAGTGGCAG -3'	5'- ACGGGCATCTTCAGGAGCTC -3'	405
Exon 5	5'- AGACGTGGAGTGACGATGCTC -3'	5'- TACCCAGCCAGATGCTCCATG -3'	678

Table 6-3 : Time and cost to sequence the coding regions of SERPINA1 in batch of 24 patients.

Steps	Time (min)*	Cost for reagents (CDN)
DNA extraction	210	$245
Quality control (260/280 nm ratio)	30	$5
PCR amplification	260	$70
Sequencing	240	$360
Data analysis	40	
Total	780 (13 hours)	$680 ($28.33/patient)

*Time is estimated for one laboratory technician.

Table 6-4 : Nomenclature of the 19 genetic variants identified and allele frequencies in this Canadian COPD population.

Exon	rs #	Nucleotide	Protein	Pi typing	MAF
2	rs745463238	c.113C>T	Ser14Phe		0.13%
2	rs150784949	c.171C>T	Phe33Phe		0.13%
2	rs28931570	c.187C>T	Arg39Ccys	I	0.25%
2	rs111850950	c.250G>A	Ala60Thr		0.25%
2	rs758820515	c.335C>T	Pro88Leu		0.13%
2	rs709932	c.374G>A	Arg101His	M2/M4	16.25%
2	rs20546	c.424C>T	Leu118Leu		1.88%
2	rs112030253	c.514G>T	Gly148Trp		0.25%
3	rs6647	c.710T>C	Val213Ala	M1	20.13%
3	New	c.738G>A	Lys222Lys		0.13%
3	rs28929470	c.739C>T	Arg223Cys	F	0.13%
3	rs759736224	c.835C>A	Pro255Thr		0.13%
3	rs121912714	c.839A>T	Asp256Val	P Lowell	0.25%
3	rs17580	c.863A>T	Glu264Val	S	6.38%
4	rs141620200	c.922G>T	Ala284Ser		0.13%
4	rs201788603	c.1061C>T	Ser330Phe		0.13%
5	rs28929474	c.1096G>A	Glu342Lys	Z	2.38%
5	rs13170	c.1188C>T	Phe372Phe		0.25%
5	rs1303	c.1200A>C	Glu376Asp	M2/M3	25.00%

MAF, minor allele frequency.

Figures

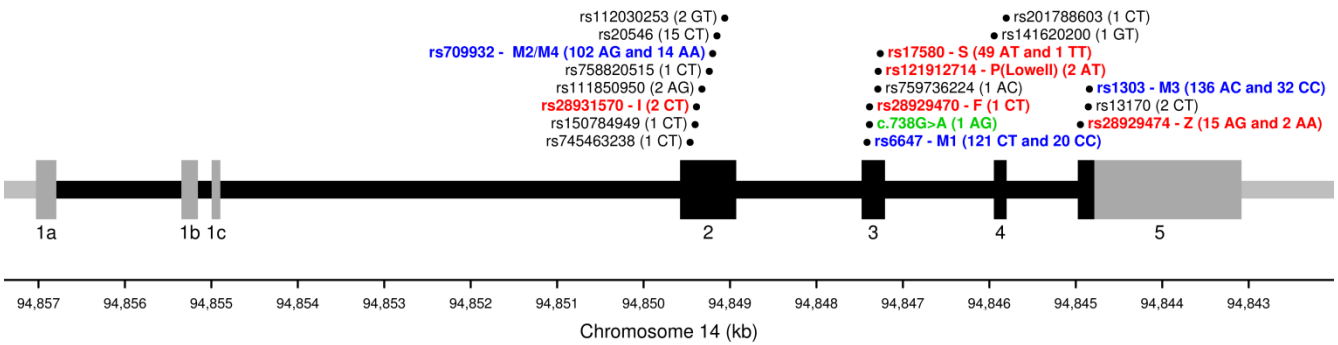

Figure 6-1 : The exon-intron structure of the SERPINA1 gene and the localization of the identified genetic variants.

Genetic variants are illustrated with their rs numbers if available with genotyping counts in parentheses for 400 individuals. Newly identified variants are illustrated in green and named based on standard gene mutation nomenclature[15]. Variants causing conventional Pi typing (M, S, Z, I, P and F) are illustrated in blue and red for normal and deficient alleles, respectively. The coding exons are shown in black and the untranslated regions in grey.

Pi typing	P	I	F	Z	S	M4	M3	M2	M1	M1*	
M1*	0	1	1	2	13	1	17	13	62	16	
M1	2	0	0	8	20	13	35	48	88		
M2	0	1	0	3	9	4	13	10			
M3	0	0	0	1	5	0	9				
M4	0	0	0	0	1	0					
S	0	0	0	1	1						
Z	0	0	0	2							
F	0	0	0								
I	0	0									
P	0										400

Figure 6-2 : Number of patients identified in any pairwise combinations of normal and deficient alleles annotated based on the legacy nomenclature.

The color of the squares illustrates the expected level of deficiency on a black and white scale where white indicates no deficiency and black indicates severe deficiency. *M1 (Ala^{213}).

CHAPITRE 7

Discussion, conclusion et perspectives

La maladie pulmonaire obstructive chronique est une maladie chronique des voies respiratoires se caractérisant par une réduction lente et progressive des fonctions respiratoires. Actuellement, aucun traitement n'existe pour renverser cette pathologie. L'identification de nouvelles cibles thérapeutiques est donc une priorité afin de diminuer les coûts sociaux-économiques de la MPOC.

Cette thèse a pour but général d'approfondir nos connaissances sur la génétique de la MPOC pour en comprendre les bases moléculaires afin de contribuer au développement d'un traitement curatif. Pour accomplir cet objectif, quatre études sur différents aspects de la maladie ont été réalisées. La première de ces études s'est attardée à identifier des gènes et des voies biologiques impliqués dans le développement de l'obstruction respiratoire (**Chapitre 3**). L'étude suivante a déterminé des gènes possiblement causaux pour différents phénotypes pulmonaires dans le complexe majeur d'histocompatibilité, une région clé du génome humain pour le système immunitaire (**Chapitre 4**). La troisième étude, un projet international regroupant 26 cohortes sur la MPOC, a identifié de nouvelles régions chromosomiques de susceptibilité pour la maladie (**Chapitre 5**). Finalement, la dernière étude avait pour objectif de vérifier si le séquençage de l'ADN peut être employé comme outil de diagnostic pour le déficit en alpha-1 antitrypsine (**Chapitre 6**). Bien que chaque article ait sa propre discussion et conclusion, les résultats principaux ainsi que plusieurs des gènes identifiés sont décrits en détail dans les pages suivantes.

7.1 Obstruction respiratoire

Mes premiers travaux de doctorat visaient à mieux comprendre les mécanismes impliqués dans le développement de l'obstruction respiratoire. Les études pangénomiques précédentes ont été fructueuses dans l'identification de régions chromosomiques et de variations génétiques[90,327–332]. Par contre, ces études n'ont pu identifier avec précision le ou les gènes causaux ni les mécanismes sous-jacents. Pour mieux comprendre les bases génétiques du développement de l'obstruction respiratoire, nous avons employé les données du consortium d'eQTL pulmonaire afin d'identifier des variations génétiques influençant l'expression de gènes, qui à leur tour sont associés à un ou des phénotypes respiratoires.

7.1.1 Modèle de causalité

La première étape de ce projet fut l'identification de modèles de causalité (*causality models*). Trois modèles différents ont été analysés lors de cette étude (causal, réactif, indépendant). Nos analyses ont détecté 169 triplets causaux (SNP → expression → phénotype) pour le VEMS% et 80 pour le ratio VEMS/CVF. Aucun triplet n'a été identifié pour les deux autres modèles, ni pour la MPOC comme trait dichotomique. Parmi les 249 triplets significatifs, 13 gènes avaient des fonctions biologiques pertinentes; 8 pour le VEMS prédit (*BCL2L1*, *CADM2*, *CST3*, *CSTA*, *FLCN*, *NCR3*, *PPARGC1A* et *TNFRSF10B*) et 5 pour le ratio (*CD22*, *DEPDC6*, *GSTO2*, *MUC22* et *SPINK5*). La fonction des gènes *CD22*, *CST3*, *GSTO2* et *PPARGC1A* ainsi que leurs mécanismes associés sont décrits en détail dans l'article. Pour ce qui est des 9 autres gènes, une description complète de leurs rôles dans le développement de l'obstruction respiratoire est présentée dans les paragraphes suivants.

BCL2L1 (*BCL2 like 1*) code pour une protéine impliquée dans la régulation de l'apoptose. Une étude parue en 2010[410] a démontré que l'extrait de fumée de cigarette augmentait l'expression de cette protéine chez les cellules dendritiques et augmentait également la survie de ces cellules. Une étude publiée en 2012 par Zhang *et al*.[411] a remarqué un phénomène semblable chez les neutrophiles. Les neutrophiles sanguins de sujets MPOC avaient une expression supérieure de *BCL2L1* comparativement aux cellules de sujets sains. De plus, l'expression de ce gène corrélait négativement avec le VEMS prédit et le ratio VEMS/CVF. Dans notre étude, une expression supérieure de ce gène était associée à un VEMS prédit inférieur, ce qui correspond aux résultats de Zhang *et al*.

CADM2 (*cell adhesion molecule 2*) est impliqué dans l'adhésion cellulaire. Ce gène semble jouer un rôle important dans la prise de décision et les fonctions cognitives[412,413] et serait impliqué dans les habitudes de consommation du cannabis[414]. Ces résultats permettent de croire que *CADM2* aurait un impact sur les fonctions respiratoires en jouant un rôle dans les habitudes de consommation des produits du tabac.

CSTA (*cystatin A*), tout comme *CST3*, est une antiprotéase à cystéine (*cysteine antiproteases*). L'expression de ce gène est augmentée chez les fumeurs et est supérieure chez ceux atteints de MPOC[415]. Tout comme *CST3*, le VEMS prédit est inférieur lorsque *CSTA* est surexprimé. Ces résultats semblent contradictoires à ceux attendus puisqu'une augmentation de *CSTA* devrait protéger les poumons contre les nombreuses protéases présentes dans les poumons de sujets MPOC. Son rôle dans le développement de l'obstruction respiratoire pourrait se faire par son effet sur les kératinocytes, des cellules impliquées dans l'épithélium. Une étude publiée en 2011[416] a démontré que des mutations présentes dans *CSTA* avaient un impact sur les interactions cellule-cellule des kératinocytes. Des études antérieures ont démontré l'implication des kératinocytes dans les poumons (développement, inflammation, réparation)[417]. De plus, le gène *FGF7* (*fibroblast growth factor 7*), aussi connu sous le nom de facteur de croissance des kératinocytes (*keratinocyte growth factor*), est impliqué dans la réparation cellulaire dans les poumons[417–419]. *CSTA* pourrait donc influencer les fonctions respiratoires non par son rôle dans la balance protéase-antiprotéase, mais par celui sur les kératinocytes.

FLCN (*folliculin*) est un gène dont la fonction précise est toujours inconnue, mais celui-ci est probablement un suppresseur de tumeur. Actuellement, ce gène est principalement connu pour son implication dans le syndrome de Birt-Hogg-Dubé[420,421], une maladie génétique rare caractérisée par des pneumothorax et des kystes pulmonaires. En 2008, une étude n'a identifié aucun lien entre ce gène et le développement de la MPOC sévères et précoces chez 152 patients[422]. Ce gène pourrait influencer le développement de pneumothorax chez les patients MPOC, une condition plus fréquente chez ces individus[423]. Il serait donc intéressant de réaliser des études d'association pour ce gène dans une cohorte de sujets MPOC avec une fréquence élevée de pneumothorax.

NCR3 (*natural cytotoxicity triggering receptor 3*) produit une protéine qui augmente l'efficacité des cellules NK activées. Deux études ont établi un lien entre des variations génétiques localisées près de ce gène et les fonctions respiratoires[339,424]. Dans nos résultats, une expression réduite de ce gène était liée à un VEMS prédit inférieur. Dans le cancer du poumon, une expression plus faible de *NCR3* dans les cellules mononuclées du sang périphérique corrèle à une survie diminuée[425]. Des études

supplémentaires devront être réalisées afin de vérifier si une corrélation semblable à celle du cancer du poumon est présente chez les patients MPOC et déterminer les mécanismes impliqués, car ceux-ci sont toujours inconnus.

TNFRSF10B (*TNF receptor superfamily member 10b*) produit un récepteur qui lorsque activé, provoque l'apoptose de la cellule[426–428]. Parmi les ligands de cette protéine on retrouve TNF-α, un gène associé à plusieurs reprises à la MPOC[8]. Une étude parue en 2009[429] a démontré que chez des cellules de muscles lisses de patients asthmatiques, l'expression de *TNFRSF10B* (*TRAILR2*) est augmentée. Morissette *et al*. ont démontré en 2008[430] que les niveaux protéiques de ce récepteur étaient augmentés dans le parenchyme pulmonaire de patients emphysémateux. Wu *et al*.[431] ont également observé des niveaux supérieurs de cette protéine dans le sérum sanguin de patients MPOC. De plus, les niveaux corrélaient négativement avec les fonctions respiratoires (VEMS prédit, ratio VEMS/CVF) et corrélaient également avec l'inflammation. Dans notre étude, une expression supérieure de ce gène était associée à un VEMS prédit plus faible, ce qui concorde avec la littérature.

DEPDC6 (*DEP domain containing MTOR interacting protein*) est un régulateur négatif de mTORC1 et mTORC2, des complexes protéiques impliqués dans plusieurs mécanismes cellulaires. Chez le fétus, ces deux molécules sont impliquées dans le développement des poumons[432]. Outre ce rôle, mTORC1 jouerait un rôle protecteur dans les poumons[433]. Fait intéressant, le gène *FLCN*, que nous avons identifié dans cette étude, serait un régulateur de mTORC1[434]. Dans nos résultats, une expression élevée de *DEPDC6* est reliée à un ratio VEMS/CVF diminué. Une expression supérieure pourrait donc provoquer l'inhibition de mTORC1 et mTORC2, et donc rendre le poumon susceptible aux dommages cellulaires.

MUC22 (*mucin 22*) fait partie de la famille des mucines, une famille de gènes impliquée dans la composition de mucus. Ce gène fut associé à l'asthme[435] et à la panbronchiolite diffuse[436], deux pathologies pulmonaires importantes. *MUC22* est surexprimé par les cellules épithéliales lors d'infections virales[437]. Un ratio VEMS/CVF moindre était associé à une expression supérieure du gène dans nos résultats. Ces données

permettent de croire que *MUC22* pourrait participer à la surproduction de mucus observée chez certains patients MPOC, plus spécifiquement ceux avec bronchite chronique.

SPINK5 (*serine peptidase inhibitor, Kazal type 5*), aussi connu sous le nom de *LEKTI* (*lympho-epithelial Kazaltype inhibitor*), est un gène codant un inhibiteur des protéases. Des variations génétiques dans ce gène ont été associées au Syndrome de Netherton, une maladie cutanée caractérisée par de l'inflammation et des allergies[438,439]. Une étude parue en 2012[440] a démontré que les cellules épithéliales qui surexprimaient le gène produisaient une plus grande quantité de facteurs proinflammatoires (IL-6, IL-8). Dans notre étude, une surexpression de ce gène dans le tissu pulmonaire était associée à un ratio VEMS/CVF inférieur. Ce gène pourrait donc influencer le développement de l'obstruction respiratoire en contribuant à l'inflammation chronique.

7.1.2 Voies biologiques

Les gènes identifiés furent par la suite employés pour identifier des voies biologiques enrichies. Plusieurs voies associées à l'apoptose, la balance oxydant-antioxydant et la régulation du système immunitaire étaient enrichies dans nos résultats. La balance oxydant-antioxydant et l'apoptose sont bien connues dans la MPOC. Tel que mentionné à la section 1.1.5.1.1, une trop grande présence de protéases dans les poumons résulte en des dommages aux cellules épithéliales et endothéliales pulmonaires. Ces lésions provoquent l'apoptose de ces cellules et participent à l'élargissement des alvéoles. Les gènes identifiés dans ce projet pourraient expliquer pourquoi l'apoptose est plus fréquente chez les cellules endothéliales de fumeurs MPOC comparativement à des fumeurs non-MPOC[441]. La réponse immunitaire adaptative est en partie responsable des symptômes persistants suite à l'arrêt tabagique (section 1.1.5.1.2). Une dérégulation des voies de signalisation du système immunitaire pourrait expliquer ce phénomène chronique.

7.1.3 Conclusion partielle 1

Ce projet a identifié de nombreux gènes et voies biologiques impliqués dans le développement de l'obstruction respiratoire. Ces résultats ont permis de mieux comprendre certains mécanismes précédemment impliqués (balance oxydant-antioxydant, apoptose,

système immunitaire) ainsi que d'en identifier des nouveaux (adhésion cellulaire, kératinocytes). De nouvelles études sont nécessaires afin de déterminer la fonction précise de certains gènes et de vérifier si ceux-ci participent réellement au développement de la MPOC. De plus, certains gènes pourraient être impliqués dans des sous-groupes particuliers.

7.2 Complexe majeur d'histocompatibilité

Les travaux suivants ont porté sur le complexe majeur d'histocompatibilité (*major histocompatibility complex*; MHC) et son rôle dans certaines maladies pulmonaires. Cette région, aussi connue sous le nom d'antigènes des leucocytes humains (*human leukocyte antigen*; HLA), est une région clé du génome humain pour le système immunitaire. Depuis sa découverte chez la souris en 1936[442], les variations génétiques et gènes de ce locus ont été reliés à plus d'une centaine de pathologies différentes[398] (e.g. cancer du poumon[376,377], maladie cœliaque[443,444], maladie inflammatoire chronique de l'intestin[445,446], MPOC[339,447–484], et psoriasis[485,486]). Tel que mentionné précédemment, ce locus possède une densité élevée de gènes. La région est aussi caractérisée par un déséquilibre de liaison important[399]. Ces deux observations compliquent grandement l'interprétation des résultats car elles rendent l'identification du ou des gènes causaux difficile.

Pour améliorer nos connaissances sur cet important locus, les données du consortium d'eQTL pulmonaire ont été utilisées afin d'identifier des gènes causaux pour plusieurs maladies et phénotypes pulmonaires (asthme, cancer du poumon, sévérité de la fibrose kystique, fonction respiratoire, maladie pulmonaire interstitielle). Bien que la MPOC ne soit pas le sujet principal de cette étude, les résultats obtenus ont permis d'approfondir nos connaissances sur plusieurs phénotypes associés.

7.2.1 Génome

La première étape de ce projet fut la vérification de la présence d'eQTL pulmonaires dans cette région et de comparer le nombre d'eQTL détectés avec d'autres régions du génome de même taille. Deux observations intéressantes furent effectuées lors de cette analyse. Premièrement, le complexe majeur d'histocompatibilité possède un

nombre important d'eQTL pulmonaire. Ce résultat était attendu puisque cette région chromosomique est extrêmement dense en terme de gènes (~50 gènes / Mb)[398,487] et des études antérieures avaient identifiées de nombreux eQTL dans cette région[488,489]. La deuxième observation est le nombre élevé d'eQTL sur le chromosome 19. Tout comme le MHC, le chromosome 19 comporte de nombreux gènes. Celui-ci est le chromosome avec la plus haute concentration de gènes par Mb[490]. Autre fait intéressant, deux études indépendantes ont détecté des « hotspot » d'eQTL distant sur ce chromosome[491,492]. Ces « hotspot » sont des variations génétiques impliquées dans de nombreux eQTL distant. Dans une étude de Wright *et al*. parue en 2014[492], un SNP localisé sur le chromosome 19 participait à un eQTL local et 19 eQTL distant. Dans une étude de Yang *et al*.[491], un eQTL local et 13 eQTL distant composaient le « hotspot ». Bien que notre étude ne fût pas sur ce chromosome, ces résultats permettent de croire que celui-ci pourrait comporter de nombreux eQTL distant qui seraient importants dans les pathologies pulmonaires.

7.2.2 Gènes causaux dans la région du MHC

Les analyses suivantes se sont attardées à bonifier les résultats obtenus par différents GWAS sur des phénotypes respiratoires avec des résultats significatifs localisés dans le MHC, plus précisément dans sa forme étendue (xMHC). 16 GWAS répondant à ces critères furent identifiés dans la littérature. Ceux-ci portaient sur l'asthme[400–405], le cancer du poumon[375,376,406,407], la sévérité de la fibrose kystique[408], les fonctions respiratoires[339,340,352,353] et la maladie pulmonaire interstitielle[409]. Des gènes avec des fonctions biologiques pertinentes ont été identifiés pour 4 des 5 phénotypes. Les paragraphes suivants décriront les résultats obtenus.

Pour l'asthme, 7 gènes (*AGPAT1*, *CDSN*, *HLA-DPB1*, *HLA-DPB2*, *HLA-DQB1*, *HLA-DQB2* et *TAP2*) participaient à des eQTL pulmonaires. Parmi ceux-ci, *HLA-DPB1/2*, *HLA-DQB1/2* et *TAP2* sont tous des gènes codant des protéines présentatrices d'antigènes. Les gènes HLA sont étudiés dans l'asthme depuis les années 1970[493,494]. Différents gènes HLA sont reliés à différents sous-types de la maladie. Par exemple, le gène *HLA-DRB1* est associé à l'asthme allergique, tandis que *HLA-DPB1* à l'asthme avec intolérance à l'aspirine[495]. Cette famille de gènes semble influencer l'asthme en ayant un impact sur les allergies, l'un des principaux facteurs de risque de la maladie. Les individus avec un

système immunitaire sensible sont plus à risque d'avoir une réaction allergique à des allergènes à priori inoffensif, ce qui peut alors occasionner une crise d'asthme. Le gène *AGPAT1* (*lysophosphatidic acid acyltransferase, alpha*) code une enzyme qui convertit l'acide lysophosphatidique (LPA) en acide phosphatidique (PA). Avant cette étude, ce gène n'avait jamais été associé à l'asthme ou autre maladie pulmonaire. Par contre, la voie biologique autotaxine-LPA, dont fait partie le gène, a été associée à cette maladie à quelques reprises[496–499]. Une des principales études de ce domaine a été publiée par Park *et al*. en 2013[498]. Les auteurs ont remarqué que la concentration d'autotaxine et de LPA était supérieure chez les sujets asthmatiques ayant subi une provocation avec un allergène. La LPA est connue pour accroître la contractilité du muscle lisse[499] et ce phospholipide est maintenant étudié comme possible biomarqueur pour l'asthme[497]. Nos résultats pour le gène *AGPAT1* concordent parfaitement avec ceux de la littérature puisque l'allèle de risque pour l'asthme est associé à une diminution de l'expression du gène. Une accumulation de LPA pourrait donc être impliquée dans des crises d'asthme plus sévères et plus régulières vu l'accroissement de la sensibilité du muscle lisse. Réduire les niveaux de LPA chez les sujets asthmatiques pourrait aider à diminuer les symptômes de la maladie. Le dernier gène associé à l'asthme dans cette étude se nomme *CDSN* (*corneodesmosin*). Ce gène est principalement associé à des maladies de peau telles que le psoriasis[500] et le syndrome de peau déciduale généralisée type B (*inflammatory peeling skin syndrome*) [501–503]. Le mécanisme impliqué pourrait être semblable à celui entre l'asthme et le gène responsable de produire la filaggrine (*FLG*). Palmer *et al*. ont publié un article en 2006[504] qui établissait un lien entre deux variations faux-sens présentes dans le gène *FLG* et la susceptibilité à l'asthme. Les auteurs ont remarqué que les sujets porteurs de ces mutations avaient un risque augmenté de souffrir d'asthme lorsque ceux-ci souffraient également de dermatite atopique. Chez les patients atteints de dermatite atopique, la barrière épithéliale présente à la surface des tissus est endommagée, ce qui permet aux allergènes de pénétrer dans les sous-couches et d'interagir avec les cellules présentatrices d'antigènes[505]. Cette interaction augmente les risques de souffrir d'allergies puisque l'organisme est déjà sensibilisé aux allergènes. *CDSN* pourrait donc avoir un impact sur le risque de souffrir d'asthme grâce à un mécanisme semblable à *FLG*.

Les résultats d'eQTL pour le cancer du poumon ont identifié deux nouveaux gènes candidats, soit *BTN3A2* et *ZFP57*. *BTN3A2* (*butyrophilin subfamily 3 member A2*) code pour une protéine possiblement impliquée dans la réponse immunitaire adaptative[506,507]. Peu d'informations sont actuellement disponibles sur ce gène. Une étude sur le cancer des ovaires a démontré qu'une concentration élevée de *BTN3A2* était associée à une infiltration supérieure de cellules T et à un meilleur pronostic[508]. Le même mécanisme pourrait également s'appliquer au cancer du poumon. Une étude publiée en 2008[509] a établi la même relation entre le nombre de cellules T (CD4+) et le cancer du poumon. Dans nos résultats, l'allèle de risque identifié dans le GWAS est associé à une expression réduite, ce qui corrobore les études précédentes. *ZFP57* (*ZFP57 zinc finger protein*) est un régulateur de la transcription et participe au maintien de l'empreinte parentale (*imprinting*), un phénomène épigénétique. Celui-ci pourrait jouer un rôle dans le développement du cancer du poumon en ayant un impact sur l'expression d'*IGF2* (*insulin like growth factor 2*). Tada *et al*. ont démontré que l'inactivation de *ZFP57* est associée à une suppression de l'expression d'*IGF2*[510]. Une augmentation de l'expression d'*IGF2* et une perte d'*imprinting* pour ce gène sont associées au développement du cancer du poumon et à une mortalité accrue chez les patients touchés[511–513]. Dans nos données, l'expression du gène est augmentée chez les sujets porteurs de l'allèle de risque, ce qui concorde avec les études précédentes

Pour les fonctions respiratoires, 9 gènes de susceptibilité ont été identifiés (*HCG4B*, *HLA-A*, *HLA-F*, *HLA-J*, *IFITM4P*, *MICB*, *PSORS1C1*, *ZFP57* et *ZSCAN31*). *HCG4B* (*HLA complex group 4B*) produit un ARN non codant. *HLA-A* (*major histocompatibility complex, class I, A*) est un gène impliqué dans le système immunitaire et est régulièrement testé lors de transplantations de rein et de moelle osseuse[514,515]. Dans une étude parue en 2016[516], un CNV couvrant ces deux gènes (*HCG4B* et *HLA-A*) fut associé à la MPOC et au VEMS. Le VEMS augmentait et le risque de souffrir de MPOC diminuait lorsqu'il y avait une ou aucune copie de ce CNV. De plus, la perte de copies diminuait l'expression des deux gènes. *MICB* (*MHC class I polypeptide-related sequence B*) code une protéine qui est un ligand pour le récepteur *NKG2D* de type II. La fonction du récepteur est de rendre les cellules susceptibles à la lyse par les cellules NK, lors de l'exposition à la fumée de cigarette[517] ou à une infection[518]. Dans notre étude, l'allèle de risque pour ce SNP

provoquait une augmentation de l'expression du gène. Une expression continue de *MICB* pourrait prédisposer les individus à une réponse inflammatoire exagérée lors d'exposition aux produits du tabac ou lors d'une infection. Cette réponse aggravée pourrait occasionner des effets délétères au tissu pulmonaire. *ZFP57*, relié au cancer du poumon dans notre étude, contient une variation génétique associée au tabagisme et aux fonctions respiratoires[353]. Dans cette étude, la variation en question fut associée au ratio VEMS/CVF en relation avec le tabagisme (déjà fumé vs jamais fumé), mais l'effet du SNP sur ces deux phénotypes n'est pas mentionné. Dans une étude de 2006[519], Silva *et al*. ont remarqué que les souris déficientes en *IGF2* (*Igf2*$^{-/-}$), gène régulé par *ZFP57*, avaient un retard de croissance pulmonaire à la fin de la gestation. Un mécanisme semblable pourrait être présent chez l'humain, ce qui se traduirait par des volumes pulmonaires plus faibles tout au long de la vie. Pour les derniers gènes (*HLA-F*, *HLA-J*, *IFITM4P*, *PSORS1C1*, *ZSCAN31*), il est plus difficile d'établir le lien entre leurs fonctions biologiques et les fonctions respiratoires.

Les maladies pulmonaires interstitielles sont un groupe de maladies rares caractérisé par de l'inflammation et de la fibrose pulmonaire[520]. Dans ce projet, l'allèle de risque était associé à des niveaux d'ARNm supérieurs de *AGPAT1*, le même gène que dans l'asthme. Cependant, l'effet de l'allèle de risque sur les niveaux d'expression était inversé entre les deux pathologies. Des niveaux supérieurs d'autotaxine et de LPA ont précédemment été observés dans des lavages broncho-alvéolaires[521], poumons[522] et condensats de l'air expiré (CAE)[523] de sujets souffrant de maladies pulmonaires interstitielles. En plus d'avoir démontré des niveaux supérieurs de LPA dans les lavages broncho-alvéolaires d'humains et de modèles murins, Tager *et al*. ont également remarqué que l'inhibition du récepteur LPA_1 protègerait de la maladie[521]. Dans notre jeu de données, l'allèle de risque était associé à une expression supérieure du gène *AGPAT1*, ce qui devrait se traduire par une diminution de la quantité de LPA. Nos résultats semblent donc être en contradiction avec ceux de la littérature qui pointent vers une accumulation de LPA. Des études supplémentaires devront être réalisées afin de répliquer nos résultats et de vérifier si un mécanisme connexe pourrait expliquer cette disparité.

La sévérité de la fibrose kystique est le seul phénotype analysé qui ne comportait aucun eQTL pulmonaire dans la région du MHC. Une explication possible est la présence d'eQTL conditionnel pour cette maladie. En effet, certains eQTL sont détectable seulement lorsqu'un ou des stimuli spécifiques sont présents[524,525]. Réaliser de nouveau cette étude en utilisant des poumons de patients souffrant de fibrose kystique permettrait peut-être d'identifier un ou des gènes causaux dans cette région.

7.2.3 Conclusion partielle 2

Cette étude a permis d'approfondir nos connaissances sur de multiples phénotypes associés à la MPOC. De possibles gènes causaux pour l'asthme, le cancer du poumon, les fonctions respiratoires et les maladies pulmonaires interstitielles furent identifiés. Ces nouvelles cibles permettent de mieux comprendre comment le MHC influence ces différents phénotypes et maladies pulmonaires. De nouvelles études sont maintenant nécessaires pour confirmer le rôle de ces gènes et déterminer avec plus de précision les mécanismes impliqués.

7.3 Régions de susceptibilité

Le troisième article de cette thèse porta sur l'identification de nouveaux gènes de susceptibilité. Lors de la dernière décennie, un grand nombre d'études pangénomiques a été réalisée afin de comprendre la composant génétique de plusieurs maladies incluant la MPOC. Les six premières études sur ce sujet ont identifié 10 régions de susceptibilité[90,327–332]. Par contre, ces loci n'expliquent qu'une faible proportion du risque de développer la maladie. La réalisation de nouveaux GWAS avec des échantillons beaucoup plus importants est donc nécessaire pour identifier de nouveaux loci. Cet objectif a mené à la création de l'*International COPD Genetics Consortium* (ICGC), un regroupement de 26 cohortes provenant de huit pays, afin de réaliser le plus important GWAS sur la MPOC.

7.3.1 Méta-analyse

Dans le GWAS de ICGC, 22 régions chromosomiques ont atteint le seuil de significativité ($P < 5x10^{-8}$) lors de la méta-analyse. Parmi les 22 loci identifiés, quatre

(*EEFSEC*, *DSP*, *MTCL1*, *SFTPD*) régions jamais associées à la MPOC ni aux fonctions respiratoires furent identifiées.

DSP (*desmoplakin*), situé sur le chromosome 6p24.3, code une protéine essentielle des desmosomes, des structures impliquées dans les liaisons entre les cellules. Ces dernières sont importantes pour maintenir l'intégrité des tissus exposés à un stress mécanique important[526]. De nombreuses études ont démontré que des mutations dans ce gène provoquent une perte d'intégrité du tissu épithélial[527–531]. Chez les souris, des expériences ont déterminé qu'une expression supérieure de la forme sauvage du gène ne comportait aucun risque[532]. Par contre, les souris avec des mutations pathologiques et qui exprimaient à la hausse le gène, développaient de multiples problèmes cardiaques. Ce gène fut également associé à la fibrose pulmonaire dans un GWAS publié en 2013[409]. En plus d'effectuer une étude pangénomique, les auteurs de l'étude ont analysé l'expression du gène dans 194 poumons humains (100 patients atteints de fibrose pulmonaire idiopathique et 94 contrôles). Ils ont remarqué 1) que l'expression du gène était supérieure chez les cas et 2) que l'allèle de risque pour la maladie était associé à des niveaux d'expression inférieur. Les auteurs n'ont pu expliquer ces résultats qui semblent se contredire. Dans notre étude, l'allèle de risque pour ce gène était associé à une augmentation de l'expression. Un mécanisme semblable à celui du cœur ou de la fibrose pulmonaire pourrait participer au développement de la MPOC. Il serait intéressant de savoir si ce gène est associé à l'emphysème ou à la bronchite chronique plus spécifiquement.

EEFSEC (*eukaryotic elongation factor, selenocysteine-tRNA specific*) est un gène localisé sur le chromosome 3q21.3 et peu connu dans la littérature. Il serait impliqué dans l'incorporation de la sélénocystéine, un acide aminé non standard, dans les sélénoprotéines. De nombreuses protéines de cette famille seraient des antioxydants[533–535]. Celles-ci pourraient donc diminuer ou prévenir le stress oxydant, l'un des processus impliqués dans le développement de la MPOC. Tel que mentionné dans l'article, *EEFSEC* est un gène paralogue de *TUFM*, un gène récemment lié à la MPOC[536–538]. La combinaison de tous ces éléments pointe vers une implication possible de *EEFSEC* dans le développement de la MPOC.

MTCL1 (*microtubule crosslinking factor 1*), mieux connu sous le nom de *SOGA2*, est impliqué dans la stabilisation des microtubules dans les cellules épithéliales[539,540]. Ce gène, présent sur le chromosome 18p11.22, est très peu connu dans la littérature. Considérant son rôle dans les cellules épithéliales, une dérégulation de ce gène pourrait donc faciliter le développement de la maladie. Cependant, les informations disponibles actuellement sont insuffisantes pour postuler sur les mécanismes impliqués.

La protéine codée par *SFTPD* (*surfactant protein D*) est impliquée dans le système immunitaire inné et dans le métabolisme du surfactant. Chez les souris *Sftpd*$^{-/-}$, l'absence de cette protéine provoque le développement d'emphysème en bas âge[541,542]. Chez l'humain, cette protéine protègerait de la MPOC en diminuant la réponse immunitaire et la production d'oxydants tout en augmentant le nettoyage des cellules mortes[543]. Il est possible qu'un phénomène semblable à celui des souris soit présent chez certains patients MPOC.

7.3.2 Corrélation génétique

Suite à la méta-analyse, nos résultats furent comparés à ceux de GWAS réalisés sur l'asthme, la densité minérale osseuse, la fibrose pulmonaire, l'indice de masse corporelle, la maladie coronarienne, le tabagisme, la taille, le VEMS et le ratio VEMS/CVF pour identifier des corrélations génétiques entre la MPOC et ces phénotypes associés. Ces analyses ont permis d'identifier des corrélations significatives entre la MPOC et l'asthme, le nombre de cigarette consommée par jour, le statut tabagique, la taille, le VEMS et le ratio VEMS/CVF. Ces résultats démontrent une fois de plus le lien important entre la MPOC et les fonctions respiratoires. La corrélation significative avec l'asthme démontre que ces deux maladies partagent des facteurs de risque génétique, ce qui vient appuyer la théorie du syndrome de chevauchement asthme-MPOC.

7.3.3 Conclusion partielle 3

Suite à la publication de cette étude, le nombre de régions associées à la MPOC est passé de 10 à 24. Parmi ces 14 nouvelles régions, quatre n'avaient jamais été identifiés dans des GWAS sur les fonctions respiratoires. Celles-ci comportaient des gènes avec des fonctions biologiques pertinentes pour le développement de la MPOC. Par contre, des

études supplémentaires devront être réalisées afin de confirmer et mieux comprendre leurs fonctions chez l'humain et leurs relations avec la maladie. De plus, il serait important de vérifier si ces gènes sont associés à l'emphysème ou la bronchite chronique plus spécifiquement.

7.4 Déficit en alpha-1 antitrypsine

Le projet final de cette thèse consistait à faire le lien entre la recherche fondamentale et la clinique. Le déficit en alpha-1 antitrypsine représente de 1 à 5% de tous les cas de MPOC et sa fréquence chez ces patients est sous-estimée. Cette situation est causée par de nombreux facteurs dont une méconnaissance de la maladie chez les médecins, des symptômes non spécifiques et des techniques de diagnostic lentes et dispendieuses. Afin d'améliorer et d'accélérer le diagnostic de la maladie, nous avons collaboré avec l'étude *CanCOLD* pour 1) déterminer les variations génétiques du gène *SERPINA1* présentes chez une population de sujets MPOC canadiens et 2) vérifier si le séquençage de l'ADN peut être employé comme outil de diagnostic.

7.4.1 Séquençage

Le séquençage des exons codants du gène *SERPINA1* a révélé la présence de 19 variations génétiques chez les 400 sujets de l'étude. Huit des 19 variations participent à 10 formes différentes de la protéine dont 5 de type Normal (M1, $M1_{Ala213}$, M2, M3, M4) et 5 Déficient (Z, S, F, I, P). De plus, 10 SNPs ont des effets inconnus sur la protéine et une nouvelle mutation fut également détectée. Les variations impliquées dans les formes Z (rs28929474) et S (rs17580) étaient présentes chez 4.25 et 12.5 % de nos sujets respectivement. Ces fréquences sont supérieures à celles observées dans la population européenne, ce qui n'est pas surprenant puisque nous étudions une cohorte de sujets MPOC.

Dans notre étude, trois sujets (0.75%) étaient considérés déficients. Ce pourcentage est inférieur à celui de la littérature qui est de 1 à 5%. Plusieurs raisons peuvent expliquer cette situation. Tout d'abord, notre étude comporte principalement des sujets de stades précoces. Puisque les personnes atteintes du déficit développent un emphysème précoce et plus sévère, il n'est pas surprenant que la fréquence de la maladie soit moindre dans nos

sujets. Reproduire cette étude mais chez 400 sujets de stades avancés produirait probablement des résultats différents. Le groupe ethnique pourrait également expliquer une partie de la différence. Tel que mentionné à la section 1.2.2.1, la fréquence de la maladie varie grandement entre les groupes ethniques et ce projet est composé de villes multiethniques.

7.4.2 Temps et coûts

Dans les recommandations des sociétés de thoracologies canadienne, américaine et européenne, le coût du séquençage était l'un des arguments pour expliquer sa non utilisation en clinique[45,220]. Notre étude a démontré que ce n'était plus le cas (~30$ par patient). Les technologies du séquençage de l'ADN évoluent à un rythme fulgurant, ce qui se traduit par une baisse des coûts tout aussi rapide. De plus, le temps nécessaire afin de réaliser ces analyses est court (13 heures de travail pour 24 patients), ce qui constitue un avantage comparativement aux techniques actuelles.

7.4.3 Conclusion partielle 4

Ce projet a démontré que le séquençage de l'ADN possède tous les prérequis pour être employé en clinique (précis, rapide, peu dispendieux). L'utilisation de cette technologie améliorerait de beaucoup le diagnostic de cette maladie génétique. Actuellement, la disponibilité des équipements et du personnel qualifié sont deux facteurs qui peuvent empêcher l'établissement de cette technique en clinique.

7.5 Conclusion et perspectives

La MPOC est une maladie courante avec de graves conséquences sur la santé des patients. Celle-ci sera la troisième cause de mortalité à l'échelle planétaire d'ici 2030. Actuellement, aucun traitement n'existe pour cette maladie et en déterminer les causes moléculaires est maintenant une priorité afin d'accélérer le développement de thérapies curatives. Cette thèse s'est principalement intéressée à mieux caractériser la contribution génétique au développement de la MPOC. D'une part, les travaux réalisés dans le cadre de cette thèse ont permis d'identifier de nombreux gènes et régions chromosomiques impliqués dans le développement de la maladie et de plusieurs phénotypes pulmonaires associés (asthme, cancer du poumon, fonctions respiratoires, pneumopathie interstitielle

idiopathique). D’autre part, ces travaux ont également permis de confirmer l’utilité du séquençage de l’ADN comme outil de diagnostic pour le déficit en alpha-1 antitrypsine.

Les traitements actuellement employés en clinique ont pour fonction de diminuer les symptômes de la MPOC, et non de traiter la maladie. Les résultats de cette thèse sont des cibles privilégiées pour de futures études sur cette maladie, tant au niveau fondamental que pour le développement de cibles thérapeutiques. Il est primordial pour ces études d’employer des sujets extrêmement bien caractérisés afin de déterminer si les gènes sont associés ou non à des sous-phénotypes cliniques précis (bronchite chronique, emphysème, syndrome de chevauchement asthme-MPOC). La découverte de gènes associés à des sous-phénotypes précis permettrait de mieux comprendre pourquoi les symptômes sont différents entre les patients et d’un jour personnaliser les traitements.

Références

1. World Health Organization. Projections of mortality and causes of death, 2015 and 2030. 2013.

2. Statistique Canada: Espérance de vie et décès. 2016.

3. Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med. 2007;176(6):532–55.

4. Hubert HB, Fabsitz RR, Feinleib M, Gwinn C. Genetic and Environmental Influences on Pulmonary Function in Adult Twins. Am Rev Respir Dis. 1982 Apr 1;125(4):409–15.

5. Lewitter FI, Tager IB, McGue M, Tishler PV, Speizer FE. Genetic and environmental determinants of level of pulmonary function. Am J Epidemiol. 1984;120(4):518–30.

6. Redline S, Tishler P V, Lewitter FI, Tager IB, Munoz A, Speizer FE. Assessment of Genetic and Nongenetic Influences on Pulmonary Function. Am Rev Respir Dis. 1987 Jan 1;135(1):217–22.

7. Zhou JJ, Cho MH, Castaldi PJ, Hersh CP, Silverman EK, Laird NM. Heritability of chronic obstructive pulmonary disease and related phenotypes in smokers. Am J Respir Crit Care Med. 2013;188(8):941–7.

8. Bossé Y. Updates on the COPD gene list. Int J COPD. 2012;7:607–31.

9. Ganrot PO, Laurell CB, Eriksson S. Obstructive lung disease and trypsin inhibitors in alpha-1-antitrypsin deficiency. Scand J Clin Lab Invest. 1967;19(3):205–8.

10. Petty TL. The history of COPD. Int J Chron Obstruct Pulmon Dis. 2006;1(1):3–14.

11. Marketos SG, Ballas CN. Bronchial asthma in the medical literature of Greek antiquity. J Asthma. 1982;19(4):263–9.

12. Crellin J. Théophile Bonet (1620-1689). Am J Pathol. 1980;98(1):212.

13. Morgagni GB. De Sedibus et causis morborum per anatomem indagatis (The seats and causes of diseases investigated by anatomy). 1769. 782 p.

14. Fishman A, Elias J, Fishman J, Grippi M, Senior R, Pack A. Fishman’s Pulmonary Diseases and Disorders, Fourth Edition. 2008. 2948 p.

15. Baillie M. The morbid anatomy of some of the most important parts of the human

body. 1793. 360 p.

16. Baillie M. A series of engravings, accompanied with explanations, which are intended to illustrate the morbid anatomy of some of the most important parts of the human body. 1803. 396 p.

17. René Théophile Hyacinthe Laennec. De l'auscultation médiate ou Traité du diagnostic des maladies des poumons et du coeur. 1819. 490 p.

18. Badham C. An Essay on Bronchitis: With a Supplement Containing Remarks on Simple Pulmonary Abscess. 1814. 185 p.

19. Strokes W. A Treatise on the Diagnosis and Treatment of Diseases of the Chest: Diseases of the Lung and Windpipe. 1837. 360 p.

20. Petty TL. John Hutchinson's Mysterious Machine Revisited. Chest. 2014;121:219S–223S.

21. Johns DP, Walters JAE, Walters EH. Diagnosis and early detection of COPD using spirometry. J Thorac Dis. 2014;6(11):1557–69.

22. Tiffeneau R, Pinelli A. Air circulant et air captif dans l'exploration de la fonction ventilatrice pulmonaire. Paris Med. 1947;37:624–8.

23. Gaensler EA. Analysis of the ventilatory defect by timed capacity measurements. Am Rev Tuberc. 1951;64(3):256–78.

24. Fishman AP. One hundred years of chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2005;171(9):941–8.

25. Fletcher C, Gilson J, Hugh-Jones P, Scadding J. Terminology, Definitions, and Classification of Chronic Pulmonary Emphysema and Related Conditions. Thorax. 1959;14(4):286–99.

26. American Thoracic Society. Chronic bronchitis, asthma and pulmonary emphysema: a statement by the Committee on Diagnostic Standards for Nontuberculous Respiratory Diseases. Vol. 85, The American review of respiratory disease. 1962.

27. Definition and classification of chronic bronchitis for clinical and epidemiological purposes. A report to the Medical Research Council by their Committee on the Aetiology of Chronic Bronchitis. Lancet. 1965;1(7389):775–9.

28. Briscoe W, Nash E. The Slow Space in Chronic Obstructive Pulmonary Disease. Ann N Y Acad Sci. 1965;121:706–22.

29. 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.

30. Pauwels RA, Buist AS, Calverley PMA, Jenkins CR, Hurd SS, on behalf of the GOLD Committee. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2001;163:1256–76.

31. Donald P. Tashkin, Celli B, Senn S, Burkhart D, Kesten S, Menjoge S, et al. A 4-Year Trial of Tiotropium in Chronic Obstructive Pulmonary Disease. N Engl J Med. 2008;359(15):1543–54.

32. Saetta M, Turato G, Facchini FM, Corbino L, Lucchini RE, Casoni G, et al. Inflammatory cells in the bronchial glands of smokers with chronic bronchitis. Am J Respir Crit Care Med. 1997;156(5):1633–9.

33. MacNee W. Pathogenesis of Chronic Obstructive Pulmonary Disease. Clin Chest Med. 2007;28(3):479–513.

34. Di Stefano A, Caramori G, Ricciardolo FLM, Capelli A, Adcock IM, Donner CF. Cellular and molecular mechanisms in chronic obstructive pulmonary disease: an overview. Clin Exp Allergy. 2004;34(8):1156–67.

35. Hogg JC. Pathophysiology of airflow limitation in chronic obstructive pulmonary disease. Lancet. 2004;364(9435):709–21.

36. Committee on the Aetiology of Chronic Bronchitis - Medical Research Council. Definition and classification of chronic bronchitis for clinical and epidemiological purposes. Lancet. 1965;213(April):775–9.

37. Global Initiative for Chronic Obstructive Lung Disease. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. 2016.

38. Anderson AE, Foraker AG. Centrilobular emphysema and panlobular emphysema: two different diseases. Thorax. 1973;28(5):547–50.

39. Kim WD, Eidelman DH, Izquierdo JL, Ghezzo H, Saetta MP, Cosio MG. Centrilobular and panlobular emphysema in smokers. Two distinct morphologic and functional entities. Am Rev Respir Dis. 1991;144(6):1385–90.

40. Nemec SF, Bankier AA, Eisenberg RL. Lower lobe-predominant diseases of the lung. Am J Roentgenol. 2013;200(4):712–28.

41. Sharafkhaneh A, Hanania NA, Kim V. Pathogenesis of emphysema: from the bench to the bedside. Proc Am Thorac Soc. 2008;5(4):475–7.

42. Takahashi M, Fukuoka J, Nitta N, Takazakura R, Nagatani Y, Murakami Y, et al. Imaging of pulmonary emphysema: a pictorial review. Int J Chron Obstruct Pulmon

Dis. 2008;3(2):193–204.

43. Bhatia R, Fromer L. Diagnosing and treating COPD: understanding the challenges and finding solutions. Int J Gen Med. 2011;4:729.

44. Celli BR, MacNee W, Agusti A, Anzueto A, Berg B, Buist AS, et al. Standards for the diagnosis and treatment of patients with COPD: a summary of the ATS/ERS position paper. Eur Respir J. 2004;23(6):932–46.

45. Marciniuk D, Hernandez P, Balter M. Alpha-1 antitrypsin deficiency targeted testing and augmentation therapy: A Canadian Thoracic Society clinical practice guideline. Can Respir J. 2012;109–16.

46. Martinez FJ, Foster G, Curtis JL, Criner G, Weinmann G, Fishman A, et al. Predictors of mortality in patients with emphysema and severe airflow obstruction. Am J Respir Crit Care Med. 2006;173(12):1326–34.

47. Fletcher C, Peto R. The natural history of chronic airflow obstruction. Br Med J. 1977;1(6077):1645–8.

48. O'Donnell DE, Banzett RB, Carrieri-Kohlman V, Casaburi R, Davenport PW, Gandevia SC, et al. Pathophysiology of dyspnea in chronic obstructive pulmonary disease: a roundtable. Proc Am Thorac Soc. 2007;4(2):145–68.

49. Hill AT, Bayley D, Stockley RA. The interrelationship of sputum inflammatory markers in patients with chronic bronchitis. Am J Respir Crit Care Med. 1999;160(3):893–8.

50. Hurst JR, Vestbo J, Anzueto A, Locantore N, Müllerova H, Tal-Singer R, et al. Susceptibility to Exacerbation in Chronic Obstructive Pulmonary Disease. N Engl J Med. 2010;363(12):1128–38.

51. Monsó E, Rosell a, Bonet G, Manterola J, Cardona PJ, Ruiz J, et al. Risk factors for lower airway bacterial colonization in chronic bronchitis. Eur Respir J. 1999;13(2):338–42.

52. Monsó E, Ruiz J, Rosell A, Manterola J, Fiz J, Morera J, et al. Bacterial infection in chronic obstructive pulmonary disease. A study of stable and exacerbated outpatients using the protected specimen brush. Am Rev Respir Dis. 1992;146(4):1067–83.

53. Sethi S, Murphy TF. Infection in the pathogenesis and course of chronic obstructive pulmonary disease. N Engl J Med. 2008;359(22):2355–65.

54. Sethi S, Wrona C, Grant BJB, Murphy TF. Strain-specific immune response to Haemophilus influenzae in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2004;169(4):448–53.

55. Soler N, Torres A, Ewig S, Gonzalez J, Celis R, El-ebiary M, et al. Bronchial

Microbial Patterns in Severe Exacerbations of Chronic Obstructive Pulmonary Disease (COPD) Requiring Mechanical Ventilation. Am J Respir Crit Care Med. 1998;157(16):1498–505.

56. Sint T, Donohue JF, Ghio AJ. Ambient air pollution particles and the acute exacerbation of chronic obstructive pulmonary disease. Inhal Toxicol. 2008;20(1):25–9.

57. Ling SH, van Eeden SF. Particulate matter air pollution exposure: role in the development and exacerbation of chronic obstructive pulmonary disease. Int J Chron Obstruct Pulmon Dis. 2009;4:233–43.

58. Donaldson GC, R Seemungal TA, Bhowmik A, Wedzicha JA. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. Thorax. 2002;57:847–52.

59. Kanner RE, Anthonisen NR, Connett JE. Lower respiratory illnesses promote FEV1 decline in current smokers but not ex-smokers with mild chronic obstructive pulmonary disease: Results from the Lung Health Study. Am J Respir Crit Care Med. 2001;164(3):358–64.

60. Rascon-Aguilar IE, Pamer M, Wludyka P, Cury J, Coultas D, Lambiase LR, et al. Role of gastroesophageal reflux symptoms in exacerbations of COPD. Chest. 2006;130(4):1096–101.

61. Lin YH, Tsai CL, Chien LN, Chiou HY, Jeng C. Newly diagnosed gastroesophageal reflux disease increased the risk of acute exacerbation of chronic obstructive pulmonary disease during the first year following diagnosis - a nationwide population-based cohort study. Int J Clin Pract. 2015;69(3):350–7.

62. Hunter J, Schmidt F, Jackson G. Exacerbations of COPD and symptoms of gastroesophageal reflux: a systematic review and meta-analysis. J Bras Pneumol. 2013;39(February):259–71.

63. Celli BR, Decramer M, Wedzicha JA, Wilson KC, Agusti AA, Criner GJ, et al. An official American Thoracic Society/European Respiratory Society statement: research questions in COPD. Eur Respir Rev. 2015;24(136):159–72.

64. Baty F, Putora PM, Isenring B, Blum T, Brutsche M. Comorbidities and Burden of COPD: A Population Based Case-Control Study. PLoS One. 2013 Jan;8(5):e63285.

65. Vestbo J, Hurd SS, Agustí AG, Jones PW, Vogelmeier C, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med. 2013;187(4):347–65.

66. Von Leupoldt A, Taube K, Lehmann K, Fritzsche A, Magnussen H. The impact of anxiety and depression on outcomes of pulmonary rehabilitation in patients with

COPD. Chest. 2011;140(3):730–6.

67. Vanfleteren LEGW, Spruit M a., Groenen M, Gaffron S, Van Empel VPM, Bruijnzeel PLB, et al. Clusters of comorbidities based on validated objective measurements and systemic inflammation in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2013;187(7):728–35.

68. Barr RG, Celli BR, Mannino DM, Petty T, Rennard SI, Sciurba FC, et al. Comorbidities, patient knowledge, and disease management in a national sample of patients with COPD. Am J Med. 2009;122(4):348–55.

69. Patel A, Hurst. Extrapulmonary comorbidities in chronic obstructive pulmonary disease: state of the art. Expert Rev Respir Med. 2011;5(5):647–62.

70. Sarkar M, Rajta P, Khatana J. Anemia in Chronic obstructive pulmonary disease: Prevalence, pathogenesis, and potential impact. Lung India. 2015;32(2):142.

71. Krishnan G, Grant BJ, Muti PC, Mishra A, Ochs-Balcom HM, Freudenheim JL, et al. Association between anemia and quality of life in a population sample of individuals with chronic obstructive pulmonary disease. BMC Pulm Med. 2006;6:23.

72. Barba R, Casasola GG De, Marco J, Emilio Losa J, Plaza S, Canora J, et al. Anemia in chronic obstructive pulmonary disease: a readmission prognosis factor. Curr Med Res Opin. 2012;28(4):617–22.

73. Martinez-Rivera C, Portillo K, Muñoz-Ferrer A, Martínez-Ortiz ML, Molins E, Serra P, et al. Anemia is a Mortality Predictor in Hospitalized Patients for COPD Exacerbation. COPD J Chronic Obstr Pulm Dis. 2012;9(3):243–50.

74. Incalzi RA, Corsonello A, Pedone C, Battaglia S, Paglino G, Bellia V. Chronic Renal Failure: A Neglected Comorbidity of COPD. Chest. 2010;137(4):831–7.

75. Ford ES, Croft JB, Mannino DM, Wheaton AG, Zhang X, Giles WH. COPD surveillance - United States, 1999-2011. Chest. 2013;144(1):284–305.

76. Yohannes AM, Baldwin RC, Connolly MJ. Prevalence of sub-threshold depression in elderly patients with chronic obstructive pulmonary disease. Int J Geriatr Psychiatry. 2006;(February):412–6.

77. Laforest L, Roche N, Devouassoux G, Belhassen M, Chouaid C, Ginoux M, et al. Frequency of comorbidities in chronic obstructive pulmonary disease, and impact on all-cause mortality: A population-based cohort study. Respir Med. 2016;117:33–9.

78. Lacasse Y, Rousseau L, Maltais F. Prevalence of Depressive Symptoms and Depression in Patients With Severe Oxygen-Dependent Chronic Obstructive Pulmonary Disease. J Cardiopulm Rehabil. 2001;21(2):80–6.

79. Hynninen KMJ, Breitve MH, Wiborg AB, Pallesen S, Nordhus IH. Psychological

characteristics of patients with chronic obstructive pulmonary disease: a review. J Psychosom Res. 2005;59(6):429–43.

80. Quint JK, Baghai-Ravary R, Donaldson GC, Wedzicha J a. Relationship between depression and exacerbations in COPD. Eur Respir J. 2008;32(1):53–60.

81. Xu W, Collet JP, Shapiro S, Lin Y, Yang T, Platt RW, et al. Independent effect of depression and anxiety on chronic obstructive pulmonary disease exacerbations and hospitalizations. Am J Respir Crit Care Med. 2008;178(9):913–20.

82. Kim HFS, Kunik ME, Molinari VA, Hillman SL, Lalani S, Orengo CA, et al. Functional Impairment in COPD Patients: The Impact of Anxiety and Depression. Psychosomatics. 2000;41(6):465–71.

83. Urff M, van den Berg J-WK, Uil SM, Chavannes NH, Damoiseaux RA. Depression and heart failure associated with clinical COPD questionnaire outcome in primary care COPD patients: a cross-sectional study. npj Prim Care Respir Med. 2014;24:14066.

84. Eisner MD, Blanc PD, Yelin EH, Katz PP, Sanchez G, Iribarren C, et al. Influence of anxiety on health outcomes in COPD. Thorax. 2010;65(3):229–34.

85. Hanania NA, Müllerova H, Locantore NW, Vestbo J, Watkins ML, Wouters EFM, et al. Determinants of depression in the ECLIPSE chronic obstructive pulmonary disease cohort. Am J Respir Crit Care Med. 2011;183(5):604–11.

86. Bratek A, Zawada K, Beil-Gawełczyk J, Beil S, Sozańska E, Krysta K, et al. Depressiveness, symptoms of anxiety and cognitive dysfunctions in patients with asthma and chronic obstructive pulmonary disease (COPD): possible associations with inflammation markers: a pilot study. J Neural Transm. 2014;(3):1–9.

87. Young RP, Hopkins RJ, Christmas T, Black PN, Metcalf P, Gamble GD. COPD prevalence is increased in lung cancer, independent of age, sex and smoking history. Eur Respir J. 2009 Aug;34(2):380–6.

88. Anthonisen NR, Skeans MA, Wise RA, Manfreda J, Kanner RE, Connett JE. The effects of a smoking cessation intervention on 14.5-year mortality. Ann Intern Med. 2005;143(8):615.

89. Divo M, Cote C, De Torres JP, Casanova C, Marin JM, Pinto-Plata V, et al. Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2012;186(2):155–61.

90. Pillai SG, Ge D, Zhu G, Kong X, Shianna K V, Need AC, et al. A Genome-Wide Association Study in Chronic Obstructive Pulmonary Disease (COPD): Identification of Two Major Susceptibility Loci. PLoS Genet. 2009;5(3):8.

91. Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D, et al. A

susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature. 2008 Apr 3;452(7187):633–7.

92. Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nat Genet. 2008;40(5):616–22.

93. Wilk JB, Shrine NRG, Loehr LR, Zhao JH, Manichaikul A, Lopez LM, et al. Genome-Wide Association Studies Identify CHRNA5/3 and HTR4 in the Development of Airflow Obstruction. Am J Respir Crit Care Med. 2012 Oct 1;186(7):622–32.

94. Pillai SG, Kong X, Edwards LD, Cho MH, Anderson WH, Coxson HO, et al. Loci identified by genome-wide association studies influence different disease-related phenotypes in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2010;182(12):1498–505.

95. Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature. 2008;452(7187):638–42.

96. Lagiou P, Trichopoulos D. Inflammatory biomarkers and risk of lung cancer. Journal of the National Cancer Institute. 2011;103(14):1073–5.

97. Coussens LM, Werb Z. Inflammation and cancer. Nature. 2002;420(6917):860–7.

98. Lee G, Walser TC, Dubinett SM. Chronic inflammation, chronic obstructive pulmonary disease, and lung cancer. Curr Opin Pulm Med. 2009;15(4):303–7.

99. Sporn MB, Liby KT. NRF2 and cancer: the good, the bad and the importance of context. Nat Rev Cancer. 2012;12(8):564–71.

100. Gosselink R, Troosters T, Decramer M. Peripheral muscle weakness contributes to exercise limitation in COPD. Am J Respir Crit Care Med. 1996;153(3):976–80.

101. Patel MS, Natanek SA, Stratakos G, Pascual S, Martínez-Llorens J, Disano L, et al. Vastus lateralis fiber shift is an independent predictor of mortality in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2014;190(3):350–2.

102. Janssens W, Lehouck A, Carremans C, Bouillon R, Mathieu C, Decramer M. Vitamin d beyond bones in chronic obstructive pulmonary disease; time to act. Am J Respir Crit Care Med. 2009;179(8):630–6.

103. Vogelmeier CF, Wouters EFM. Treating the systemic effects of chronic obstructive pulmonary disease. Proc Am Thorac Soc. 2011;8(4):376–9.

104. Glass DJ. Skeletal muscle hypertrophy and atrophy signaling pathways. Int J Biochem Cell Biol. 2005;37(10 SPEC. ISS.):1974–84.

105. Späte U, Schulze PC. Proinflammatory cytokines and skeletal muscle. Curr Opin Clin Nutr Metab Care. 2004;7(3):265–9.

106. Allaire J, Maltais FF, LeBlanc P, Simard P-M, Whittom FF, Doyon J-FJ-F, et al. Lipofuscin accumulation in the vastus lateralis muscle in patients with chronic obstructive pulmonary disease. Muscle Nerve. 2002;25(3):383–9.

107. Barreiro E, Gea J, Corominas JM, Hussain SNA. Nitric Oxide Synthases and Protein Oxidation in the Quadriceps Femoris of Patients with Chronic Obstructive Pulmonary Disease. Am J Respir Cell Mol Biol. 2003;29(6):771–8.

108. Hoppeler H, Kleinert E, Schlegel C, Claassen H, Howald H, Kayar S, et al. II. Morphological Adaptations of Human Skeletal Muscle to Chronic Hypoxia*. Int J Sports Med. 1990;11(S 1):S3–9.

109. England BK, Chastain JL, Mitch WE. Abnormalities in protein synthesis and degradation induced by extracellular pH in BC3H1 myocytes. Am J Physiol. 1991;260(2 Pt 1):C277–82.

110. Fiaccadori E, Del Canale S, Vitali P, Coffrini E, Ronda N, Guariglia A. Skeletal muscle energetics, acid-base equilibrium and lactate metabolism in patients with severe hypercapnia and hypoxemia. Chest. 1987;92(5):883–7.

111. Sajjan US. Susceptibility to viral infections in chronic obstructive pulmonary disease: role of epithelial cells. Curr Opin Pulm Med. 2013;19(2):125–32.

112. Hogg JC, Timens W. The pathology of chronic obstructive pulmonary disease. Annu Rev Pathol. 2009;4:435–59.

113. van Mourik Y, Bertens LCM, Cramer MJM, Lammers J-WJ, Reitsma JB, Moons KGM, et al. Unrecognized Heart Failure and Chronic Obstructive Pulmonary Disease (COPD) in Frail Elderly Detected Through a Near-Home Targeted Screening Strategy. J Am Board Fam Med. 2014;27(6):811–21.

114. Mapel DW, Dedrick D, Davis K. Trends and cardiovascular co-morbidities of COPD patients in the Veterans Administration Medical System, 1991-1999. COPD. 2005;2(1):35–41.

115. Müllerova H, Agusti A, Erqou S, Mapel DW. Cardiovascular comorbidity in COPD: Systematic literature review. Chest. 2013;144(4):1163–78.

116. Singanayagam A, Schembri S, Chalmers JD. Predictors of mortality in hospitalized adults with acute exacerbation of chronic obstructive pulmonary disease. Ann Am Thorac Soc. 2013;10(2):81–9.

117. Patel ARC, Donaldson GC, Mackay AJ, Wedzicha J a., Hurst JR. The impact of ischemic heart disease on symptoms, health status, and exacerbations in patients with COPD. Chest. 2012;141(4):851–7.

118. Chen PS, Chen LS, Fishbein MC, Lin SF, Nattel S. Role of the autonomic nervous system in atrial fibrillation: Pathophysiology and therapy. Circ Res. 2014;114(9):1500–15.

119. Mehra R, Benjamin EJ, Shahar E, Gottlieb DJ, Nawabit R, Kirchner HL, et al. Association of nocturnal arrhythmias with sleep-disordered breathing: The sleep heart health study. Am J Respir Crit Care Med. 2006;173(8):910–6.

120. McAllister D a., Maclay JD, Mills NL, Mair G, Miller J, Anderson D, et al. Arterial stiffness is independently associated with emphysema severity in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2007;176(12):1208–14.

121. Boschetto P, Beghé B, Fabbri LM, Ceconi C. Link between chronic obstructive pulmonary disease and coronary artery disease: implication for clinical practice. Respirology. 2012;17(3):422–31.

122. Mapel D. Renal and hepatobiliary dysfunction in chronic obstructive pulmonary disease. Curr Opin Pulm Med. 2014;20(2):186–93.

123. Hasegawa W, Yamauchi Y, Yasunaga H, Sunohara M, Jo T, Matsui H, et al. Factors affecting mortality following emergency admission for chronic obstructive pulmonary disease. BMC Pulm Med. 2014;14(1):151.

124. Chen C-Y, Hsu T-W, Mao SJT, Chang S-C, Yang P-C, Lee Y-C, et al. Abnormal renal resistive index in patients with mild-to-moderate chronic obstructive pulmonary disease. COPD. 2013;10(2):216–25.

125. Nangaku M. Hypoxia and tubulointerstitial injury: A final common pathway to end-stage renal failure. Nephron - Exp Nephrol. 2004;98(1):17–25.

126. Lehouck A, Boonen S, Decramer M, Janssens W. COPD, bone metabolism, and osteoporosis. Chest. 2011;139(3):648–57.

127. Graat-Verboom L, Wouters EFM, Smeen FWJM, Van Den Borne BEEM, Lunde R, Spruit M a. Current status of research on osteoporosis in COPD: A systematic review. Eur Respir J. 2009;34(1):209–18.

128. Romme E a PM, Murchison JT, Edwards LD, van Beek E, Murchison DM, Rutten EP a, et al. CT-measured bone attenuation in patients with chronic obstructive pulmonary disease: relation to clinical features and outcomes. J Bone Miner Res. 2013;28(6):1369–77.

129. Casanova C, Baudet JS, del Valle Velasco M, Martin JM, Aguirre-Jaime a., Pablo de Torres J, et al. Increased gastro-oesophageal reflux disease in patients with severe COPD. Eur Respir J. 2004;23(6):841–5.

130. Kempainen RR, Savik K, Whelan TP, Dunitz JM, Herrington CS, Billings JL. High

prevalence of proximal and distal gastroesophageal reflux disease in advanced COPD. Chest. 2007;131(6):1666–71.

131. Tsai C-L, Lin Y-H, Wang M-T, Chien L-N, Jeng C, Chian C-F, et al. Gastro-oesophageal reflux disease increases the risk of intensive care unit admittance and mechanical ventilation use among patients with chronic obstructive pulmonary disease: a nationwide population-based cohort study. Crit Care. 2015;19(1):1–11.

132. Clayton N a, Carnaby GD, Peters MJ, Ing AJ. Impaired laryngopharyngeal sensitivity in patients with COPD: The association with swallow function. Int J Speech Lang Pathol. 2014;16(6):615–23.

133. Pauwels A, Blondeau K, Dupont LJ, Sifrim D. Mechanisms of increased gastroesophageal reflux in patients with cystic fibrosis. Am J Gastroenterol. 2012;107(9):1346–53.

134. Kim J, Lee JH, Kim YYS, Kim K, Oh Y-M, Yoo KH, et al. Association between chronic obstructive pulmonary disease and gastroesophageal reflux disease: a national cross-sectional cohort study. BMC Pulm Med. 2013;13(1):51.

135. Hersh CP, Make BJ, Lynch D a, Barr R, Bowler RP, Calverley PM, et al. Non-emphysematous chronic obstructive pulmonary disease is associated with diabetes mellitus. BMC Pulm Med. 2014;14(1):164.

136. Kinney GL, Black-Shinn JL, Wan ES, Make B, Regan E, Lutz S, et al. Pulmonary function reduction in diabetes with and without chronic obstructive pulmonary disease. Diabetes Care. 2014;37(2):389–95.

137. Li LSK, Caughey GE, Johnston KN. The association between co-morbidities and physical performance in people with chronic obstructive pulmonary disease: a systematic review. Chron Respir Dis. 2014;11(1):3–13.

138. Mannino DM, Thorn D, Swensen a., Holguin F. Prevalence and outcomes of diabetes, hypertension and cardiovascular disease in COPD. Eur Respir J. 2008;32(4):962–9.

139. Koskela HO, Salonen PH, Romppanen J, Niskanen L. A history of diabetes but not hyperglycaemia during exacerbation of obstructive lung disease has impact on long-term mortality: a prospective, observational cohort study. BMJ Open. 2015;5(1):e006794.

140. García-Olmos L, Alberquilla A, Ayala V, García-Sagredo P, Morales L, Carmona M, et al. Comorbidity in patients with chronic obstructive pulmonary disease in family practice: a cross sectional study. BMC Fam Pract. 2013;14:11.

141. Kostikas K, Clemens A, Patalano F. The asthma – COPD overlap syndrome : do we really need another syndrome in the already complex matrix of airway disease ? Int J COPD. 2016;11:1297–306.

142. Postma DS, Rabe KF. The Asthma–COPD Overlap Syndrome. New Engl J Med. 2015;373(13):1241–9.

143. Silva GE, Sherrill DL, Guerra S, Barbee RA. Asthma as a risk factor for COPD in a longitudinal study. Chest. 2004;126(1):59–65.

144. Kohansal R, Martinez-Camblor P, Agustí A, Buist AS, Mannino DM, Soriano JB. The natural history of chronic airflow obstruction revisited: an analysis of the Framingham offspring cohort. Am J Respir Crit Care Med. 2009;180(1):3–10.

145. Adeloye D, Lee C, Papana A, Sridhar D, Chan KY. Global and regional estimates of COPD prevalence: Systematic review and meta–analysis. J Glob Health. 2015;5(2).

146. Statistique Canada: Maladies et état de santé. 2016.

147. Doucet M, Rochette L, Hamel D. Incidence, Prevalence, and Mortality Trends in Chronic Obstructive Pulmonary Disease over 2001 to 2011: A Public Health Point of View of the Burden. Can Respir J. 2016;2016(7518287).

148. Rycroft CE, Heyes A, Lanza L, Becker K. Epidemiology of chronic obstructive pulmonary disease : a literature review. Int J Chron Obstruct Pulmon Dis. 2012;7:457–94.

149. Khakban A, Sin DD, FitzGerald JM, Ng R, Zafari Z, McManus B, et al. Ten-Year Trends in Direct Costs of COPD. Chest. 2015;148(3):640.

150. Mittmann N, Kuramoto L, Seung SJ, Haddon JM, Bradley-Kennedy C, Fitzgerald JM. The cost of moderate and severe COPD exacerbations to the Canadian healthcare system. Respir Med. 2008;102(3):413–21.

151. Theriault L, Hermus G, Goldfarb D, Stonebridge C, Bounajm F. Cost risk analysis for chronic lung disease in Canada. Conf Board Canada. 2012;62.

152. Ford ES, Murphy LB, Khavjou O, Giles WH, Holt JB, Croft JB. Total and state-specific medical and absenteeism costs of COPD among adults aged >= 18 years in the United States for 2010 and projections through 2020. Chest. 2015;147(1):31–45.

153. Chen JC, Mannino DM. Worldwide epidemiology of chronic obstructive pulmonary disease. Curr Opin Pulm Med. 1999;5(2):93–9.

154. Løkke A, Lange P, Scharling H, Fabricius P, Vestbo J. Developing COPD: a 25 year follow up study of the general population. Thorax. 2006 Nov;61(11):935–9.

155. Davis RM, Novotny TE. 3 . Changes in Risk Factors Chronic Obstructive Pulmonary Disease 1-3. Am Rev Respir Dis. 1989;1–3.

156. Fletcher Charles, Peto Richard, Tinker Cecily SFE. The Natural History of Chronic Bronchitis and Emphysema. Ann Intern Med. 1978;88:445.

157. Lundbäck B, Lindberg A, Lindström M, Rönmark E, Jonsson AC, Jönsson E, et al. Not 15 but 50% of smokers develop COPD?--Report from the Obstructive Lung Disease in Northern Sweden Studies. Respir Med. 2003;97(2):115–22.

158. Woodruff PG, Barr RG, Bleecker E, Christenson SA, Couper D, Curtis JL, et al. Clinical Significance of Symptoms in Smokers with Preserved Pulmonary Function. N Engl J Med. 2016;374(19):1811–21.

159. Welte T, Vogelmeier C, Papi A. COPD: Early diagnosis and treatment to slow disease progression. Int J Clin Pract. 2015;69(3):336–49.

160. Silverman EK. Progress in Chronic Obstructive Pulmonary Disease Genetics. Proc Am Thorac Soc. 2006;3(5):405–8.

161. Hoffmann D, Hoffmann I, El-Bayoumy K. The less harmful cigarette: a controversial issue. a tribute to Ernst L. Wynder. Chem Res Toxicol. 2001;14(7):767–90.

162. Martin TR, Frevert CW. Innate immunity in the lungs. Proc Am Thorac Soc. 2005;2(5):403–11.

163. Carnevali S, Petruzzelli S, Longoni B, Vanacore R, Barale R, Cipollini M, et al. Cigarette smoke extract induces oxidative stress and apoptosis in human lung fibroblasts. Am J Physiol Lung Cell Mol Physiol. 2003;284(6):L955–63.

164. MacNee W. Oxidants/antioxidants and COPD. Chest. 2000;117(5 Suppl 1):303S–17S.

165. Leopold PL, O'Mahony MJ, Julie Lian X, Tilley AE, Harvey BG, Crystal RG. Smoking is associated with shortened airway Cilia. PLoS One. 2009;4(12).

166. Akira S, Uematsu S, Takeuchi O. Pathogen Recognition and Innate Immunity. Cell. 2006;124(4):783–801.

167. Tait Wojno ED, Artis D. Innate lymphoid cells: Balancing immunity, inflammation, and tissue repair in the intestine. Cell Host Microbe. 2012;12(4):445–7.

168. Abboud RT, Vimalanathan S. Pathogenesis of COPD. Part I. The role of protease-antiprotease imbalance in emphysema. Int J Tuberc lung Dis Off J Int Union against Tuberc Lung Dis. 2008;12(4):361–7.

169. Madri J, Graesser D. Cell migration in the immune system: the evolving inter-related roles of adhesion molecules and proteinases. Dev Immunol. 2000;7(2–4):103–16.

170. Pham CTN. Neutrophil serine proteases: specific regulators of inflammation. Nat Rev Immunol. 2006;6(7):541–50.

171. Parks WC. Matrix metalloproteinases in lung repair. Eur Respir J Suppl.

2003;44:36s–38s.

172. Chua F, Laurent GJ. Neutrophil Elastase: Mediator of Extracellular Matrix Destruction and Accumulation. Proc Am Thorac Soc. 2006;3(5):424–7.

173. Kawabata K, Hagio T, Matsuoka S. The role of neutrophil elastase in acute lung injury. Eur J Pharmacol. 2002;451(1):1–10.

174. Churg A, Wright JL. Proteases and emphysema. Curr Opin Pulm Med. 2005;11:153–9.

175. Macnee W. Pathogenesis of Chronic Obstructive Pulmonary Disease. Proc Am Thorac Soc. 2005;2(3):258–66.

176. Crystal RG. Alpha 1-antitrypsin deficiency, emphysema, and liver disease. Genetic basis and strategies for therapy. J Clin Invest. 1990;85(5):1343–52.

177. Park HS, Kim SR, Lee YC. Impact of oxidative stress on lung diseases. Respirology. 2009;14(1):27–38.

178. Kirkham PA, Barnes PJ. Oxidative stress in COPD. Chest. 2013;144(1):266–73.

179. Devasagayam TPA, Tilak J, Boloor KK, Sane KS, Ghaskadbi SS, Lele RD. Free radicals and antioxidants in human health: current status and future prospects. J Assoc Physicians India. 2004;52(October):794–804.

180. Valko M, Leibfritz D, Moncol J, Cronin MT, Mazur M, Telser J. Free radicals and antioxidants in normal physiological functions and human disease. IntJBiochemCell Biol. 2007;39(1):44–84.

181. Yang Y, Bazhin A V, Werner J, Karakhanova S. Reactive oxygen species in the immune system. Int Rev Immunol. 2013;32(3):249–70.

182. Faux SP, Tai T, Thorne D, Xu Y, Breheny D, Gaca M. The role of oxidative stress in the biological responses of lung epithelial cells to cigarette smoke. Biomarkers. 2009;14 Suppl 1(June):90–6.

183. Louhelainen N, Rytilä P, Haahtela T, Kinnula VL, Djukanović R. Persistence of oxidant and protease burden in the airways after smoking cessation. BMC Pulm Med. 2009;9:25.

184. Donohue JF. Ageing, smoking and oxidative stress. Thorax. 2006;61(6):461–2.

185. Hodge S, Hodge G, Holmes M, Reynolds PN. Increased airway epithelial and T-cell apoptosis in COPD remains despite smoking cessation. Eur Respir J. 2005;25(3):447–54.

186. Cosio MG, Saetta M, Agusti A. Immunologic Aspects of Chronic Obstructive Pulmonary Disease. N Engl J Med. 2009;360(23):2445–54.

187. Willemse BWM, ten Hacken NHT, Rutgers B, Lesman-Leegte IG a T, Postma DS, Timens W. Effect of 1-year smoking cessation on airway inflammation in COPD and asymptomatic smokers. Eur Respir J Off J Eur Soc Clin Respir Physiol. 2005 Nov;26(5):835–45.

188. Pouwels SD, Heijink IH, ten Hacken NHT, Vandenabeele P, Krysko D V, Nawijn MC, et al. DAMPs activating innate and adaptive immune responses in COPD. Mucosal Immunol. 2014;7(2):215–26.

189. Brusselle GG, Demoor T, Bracke KR, Brandsma CA, Timens W. Lymphoid follicles in (very) severe COPD: Beneficial or harmful? Eur Respir J. 2009;34(1):219–30.

190. Mauri C, Bosma A. Immune regulatory function of B cells. Annu Rev Immunol. 2012;30:221–41.

191. Lebien TW, Tedder TF. B lymphocytes : how they develop and function. Am Soc Hematol. 2008;112(5):1570–80.

192. Hogg JC, Chu F, Utokaparch S, Woods R, Elliott WM, Buzatu L, et al. The Nature of Small-Airway Obstruction in Chronic Obstructive Pulmonary Disease. N Engl J Med. 2004;350(26):2645–53.

193. Chung KF, Adcock IM. Multifaceted mechanisms in COPD: Inflammation, immunity, and tissue repair and destruction. Eur Respir J. 2008;31(6):1334–56.

194. Zhu J, Paul WE. CD4 T cells : fates , functions , and faults. Blood. 2009;112(5):1557–69.

195. Luckheeram RV, Zhou R, Verma AD, Xia B. CD4 +T cells: Differentiation and functions. Clin Dev Immunol. 2012;2012(925135).

196. Forsslund H, Yang M, Karimi R, Nyrén S, Grunewald J, Merikallio H, et al. Gender differences in the T-cell profiles of the airways in COPD patients associated with clinical phenotypes. Int J COPD. 2017;12:35–48.

197. Swain SL, McKinstry KK, Strutt TM. Expanding roles for $CD4^+$ T cells in immunity to viruses. Nat Rev Immunol. 2012;12(2):136–48.

198. Amsen D, Spilianakis CG, Flavell RA. How are TH1 and TH2 effector cells made? Curr Opin Immunol. 2006;21(2):153–60.

199. Löfdahl MJ, Roos-Engstrand E, Pourazar J, Bucht A, Dahlen B, Elmberger G, et al. Increased intraepithelial T-cells in stable COPD. Respir Med. 2008;102(12):1812–8.

200. Zhang J, Chu S, Zhong X, Lao Q, He Z, Liang Y. Increased expression of CD4+IL-17+ cells in the lung tissue of patients with stable chronic obstructive pulmonary disease (COPD) and smokers. Int Immunopharmacol. 2013;15(1):58–66.

201. Retamales I, Elliott WM, Meshi B, Coxson HO, Pare PD, Sciurba FC, et al. Amplification of inflammation in emphysema and its association with latent adenoviral infection. Am J Respir Crit Care Med. 2001;164(3):469–73.

202. Brusselle GG, Joos GF, Bracke KR. New insights into the immunology of chronic obstructive pulmonary disease. Lancet. 2011;378(9795):1015–26.

203. Smyth LJC, Starkey C, Vestbo J, Singh D. CD4-regulatory cells in COPD patients. Chest. 2007;132(1):156–63.

204. Freeman CM, Han MK, Martinez FJ, Murray S, Liu LX, Chensue SW, et al. Cytotoxic potential of lung CD8(+) T cells increases with chronic obstructive pulmonary disease severity and with in vitro stimulation by IL-18 or IL-15. J Immunol. 2010;184(11):6504–13.

205. Morissette MC, Parent J, Milot J. Perforin, granzyme B, and FasL expression by peripheral blood T lymphocytes in emphysema. Respir Res. 2007;8:62.

206. Andersen MH, Schrama D, Straten PT, Becker JC, Thor Straten, P. J.C. B. Cytotoxic T cells. Invest Dermatol. 2006;126(1):32–41.

207. Zhu X, Gadgil AS, Givelber R, George MP, Stoner MW, Sciurba FC, et al. Peripheral T cell functions correlate with the severity of chronic obstructive pulmonary disease. J Immunol. 2009;182(5):3270–7.

208. Chrysofakis G, Tzanakis N, Kyriakoy D, Tsiligianni I, Klimathianaki M, M N. Perforin Expression and Cytotoxic Activity of Sputum CD8 + Lymphocytes in Patients With COPD. Chest. 2004;71–6.

209. Cosio MG, Majo J, Cosio MG. Phosphodiesterase 4 inhibitors for the treatment of COPD. Chest. 2002;121(5 SUPPL.):160S–165S.

210. Saetta M, Mariani M, Panina-Bordignon P, Turato G, Buonsanti C, Baraldo S, et al. Increased expression of the chemokine receptor CXCR3 and its ligand CXCL10 in peripheral airways of smokers with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2002;165(10):1404–9.

211. Majo J, Ghezzo H, Cosio MG. Lymphocyte population and apoptosis in the lungs of smokers and their relation to emphysema. Eur Respir J. 2001;17(5):946–53.

212. Saetta M, Turato G, Maestrelli P, Mapp CE, Fabbri LM. Cellular and Structural Bases of Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2001;163:1304–9.

213. Saetta M, Baraldo S, Corbino L, Turato G, Braccioni F, Rea F, et al. CD8 + ve Cells in the Lungs of Smokers with Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 1999;160(2):711–7.

214. Marina S, Antonino DS, Graziella T, Fabrizio F, Laura C, Cristina M, et al. CD8+ T Lymphocytes in peripheral airways of smokers with chronic obstructive pulmonary disease. Am J Crit Care Med. 1998;157(4):822–826 ST–CD8+ T Lymphocytes in peripheral air.

215. O'Shaughnessy TC, Ansari TW, Barnes NC, Jeffery PK. Inflammation in bronchial biopsies of subjects with chronic bronchitis: Inverse relationship of CD8+ T lymphocytes with FEV1. Am J Respir Crit Care Med. 1997;155(3):852–7.

216. Bradi AC, Audisho N, Casey DK, Chapman KR. Alpha-1 antitrypsin deficiency in Canada: regional disparities in diagnosis and management. COPD. 2015;12 Suppl 1(December):15–21.

217. De Serres FJ. Worldwide racial and ethnic distribution of alpha1-antitrypsin deficiency: summary of an analysis of published genetic epidemiologic surveys. Chest. 2002;122(5):1818–29.

218. Lieberman J, Winter B, Sastre A. Alpha 1-antitrypsin Pi-types in 965 COPD patients. Chest. 1986;89(3):370–3.

219. Abboud R, Jung B, Mattman A. Alpha1-antitrypsin deficiency: a clinical-genetic overview. Appl Clin Genet. 2011 Mar;4:55–65.

220. American Thoracic Society/European Respiratory Society. American Thoracic Society/European Respiratory Society statement: standards for the diagnosis and management of individuals with alpha-1 antitrypsin deficiency. Am J Respir Crit Care Med. 2003;168(7):818–900.

221. Campos MA, Wanner A, Zhang G, Sandhaus RA. Trends in the diagnosis of symptomatic patients with ??1- antitrypsin deficiency between 1968 and 2003. Chest. 2005;128(3):1179–86.

222. Bornhorst JA, Greene DN, Ashwood ER, Grenache DG. alpha1-Antitrypsin phenotypes and associated serum protein concentrations in a large clinical population. Chest. 2013;143(4):1000–8.

223. Steiner SJ, Gupta SK, Croffie JM, Fitzgerald JF. Serum levels of alpha1-antitrypsin predict phenotypic expression of the alpha1-antitrypsin gene. Dig Dis Sci. 2003 Sep;48(9):1793–6.

224. Greulich T, Vogelmeier CF. Alpha-1-antitrypsin deficiency: increasing awareness and improving diagnosis. Ther Adv Respir Dis. 2015;1753465815602162.

225. Borawski J, Naumnik B, Myśliwiec M. Serum α 1 -Antitrypsin But Not Complement C3 and C4 Predicts Chronic Inflammation in Hemodialysis Patients. Ren Fail. 2003 Jan;25(4):589–93.

226. Ferrarotti I, Thun G a., Zorzetto M, Ottaviani S, Imboden M, Schindler C, et al.

Serum levels and genotype distribution of 1-antitrypsin in the general population. Thorax. 2012;67(8):669–74.

227. Prins J, Van Der Meijden BB, Kraaijenhagen RJ, Wielders JPM. Inherited chronic obstructive pulmonary disease: New selective-sequencing workup for ??1-antitrypsin deficiency identifies 2 previously unidentified null alleles. Clin Chem. 2008;54(1):101–7.

228. Postma DS, Bush A, van den Berge M. Risk factors and early origins of chronic obstructive pulmonary disease. Lancet. 2015;385(9971):899–909.

229. Vestbo J, Lange P. Natural history of COPD: Focusing on change in FEV1. Respirology. 2016;21(1):34–43.

230. Ghanbari Ghozikali M, Heibati B, Naddafi K, Kloog I, Oliveri Conti G, Polosa R, et al. Evaluation of Chronic Obstructive Pulmonary Disease (COPD) attributed to atmospheric O3, NO2, and SO2 using Air Q Model (2011-2012 year). Environ Res. 2016;144:99–105.

231. Kauffmann F, Drouet D, Lellouch J, Brille D. Occupational exposure and 12-year spirometric changes among Paris area workers. Br J Ind Med. 1982;39(3):221–32.

232. Ko FWS, Hui DSC. Air pollution and chronic obstructive pulmonary disease. Respirology. 2012;17(3):395–401.

233. Silva R, Oyarzún M, Olloquequi J. Pathogenic Mechanisms in Chronic Obstructive Pulmonary Disease Due to Biomass Smoke Exposure. Arch Bronconeumol (English Ed. 2015;51(6):285–92.

234. Barker DJ, Godfrey KM, Fall C, Osmond C, Winter PD, Shaheen SO. Relation of birth weight and childhood respiratory infection to adult lung function and death from chronic obstructive airways disease. BMJ. 1991;303(6804):671–5.

235. Stern DA, Morgan WJ, Wright AL, Guerra S, Martinez FD. Poor airway function in early infancy and lung function by age 22 years: a non-selective longitudinal cohort study. Lancet. 2007;370(9589):758–64.

236. Grevink G, Duiverman EJ, Vrijlandt EJLE, Gerritsen J, Boezen HM. Lung Function and Exercise Capacity in Young Adults Born Prematurely. Am J Respir Crit Care Med. 2006;173:890–6.

237. Saarenpää H-K, Tikanmäki M, Sipola-Leppänen M, Hovi P, Wehkalampi K, Siltanen M, et al. Lung Function in Very Low Birth Weight Adults. Pediatrics. 2015;136(4):642–50.

238. Aryal S, Diaz-Guzman E, Mannino DM. COPD and gender differences: An update. Transl Res. 2013;162(4):208–18.

239. World Health Organization. Tobacco use data by country. 2015.

240. Sin DD, Cohen SB-Z, Day A, Coxson H, Paré PD. Understanding the biological differences in susceptibility to chronic obstructive pulmonary disease between men and women. Proc Am Thorac Soc. 2007;4(8):671–4.

241. Burrows B, Bloom JW, Traver GA, Cline NG. The course and prognosis of different forms of chronic airways obstruction in a sample from the general population. New Engl J Med. 1987;317(21):1309–14.

242. De Torres JP, Casanova C, Hernández C, Abreu J, Aguirre-Jaime A, Celli BR. Gender and COPD in patients attending a pulmonary clinic. Chest. 2005;128(4):2012–6.

243. Huisman M, Kunst AE, Bopp M, Borgan J-K, Borrell C, Costa G, et al. Educational inequalities in cause-specific mortality in middle-aged and older men and women in eight western European populations. Lancet. 2005;365(9458):493–500.

244. Gershon A, Dolmage T, Stephenson A, Jackson B. Chronic Obstructive Pulmonary Disease and Socioeconomic Status: A Systematic Review. COPD J Chronic Obstr Pulm Dis. 2012;9(3):216–26.

245. Kanervisto M, Vasankari T, Laitinen T, Heliövaara M, Jousilahti P, Saarelainen S. Low socioeconomic status is associated with chronic obstructive airway diseases. Respir Med. 2011;105(8):1140–6.

246. Barker DJ, Osmond C. Childhood respiratory infection and adult chronic bronchitis in England and Wales. Br Med J (Clin Res Ed). 1986;293(6557):1271–5.

247. Burney P, Jithoo A, Kato B, Janson C, Mannino D, Nizankowska-Mogilnicka E, et al. Chronic obstructive pulmonary disease mortality and prevalence: the associations with smoking and poverty--a BOLD analysis. Thorax. 2014;69(5):465–73.

248. Taylor DH, Hasselblad V, Henley SJ, Thun MJ, Sloan FA. Benefits of Smoking Cessation for Longevity. Am J Public Health. 2002 Jun;92(6):990–6.

249. Jha P, Ramasundarahettige C, Landsman V, Rostron B, Thun M, Anderson RN, et al. 21st-century hazards of smoking and benefits of cessation in the United States. N Engl J Med. 2013;368(4):341–50.

250. Sanei F, Wilkinson T. Influenza vaccination for patients with chronic obstructive pulmonary disease: understanding immunogenicity, efficacy and effectiveness. Ther Adv Respir Dis. 2016;

251. Papi A, Luppi F, Franco F, Fabbri LM. Pathophysiology of Exacerbations of Chronic Obstructive Pulmonary Disease. Proc Am Thorac Soc. 2006;3(11):245–51.

252. O'Donnell DE. COPD exacerbations {middle dot} 3: Pathophysiology. Thorax.

2006;61(4):354–61.

253. Hsu AC-Y, Starkey MR, Hanish I, Parsons K, Haw TJ, Howland LJ, et al. Targeting PI3K-p110α Suppresses Influenza Virus Infection in Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2015;191(9):1012–23.

254. Bernard S, Ribeiro F, Maltais F, Saey D. Prescribing exercise training in pulmonary rehabilitation: A clinical experience. Rev Port Pneumol. 2014;20(2):92–100.

255. Maltais F, Bourbeau J, Shapiro S, Lacasse Y, Perrault H, Baltzan M, et al. Effects of Home-Based Pulmonary Rehabilitation in Patients with Chronic Obstructive Pulmonary Disease. Ann Intern Med. 2008;149:869–78.

256. Maltais F, Bourbeau J, Shapiro S, Lacasse Y, Perrault H, Baltzan M, et al. Home-based pulmonary rehabilitation in chronic obstructive pulmonary disease patients. Curr Opin Pulm Med. 2003;149(2):873–9.

257. Maltais F, Decramer M, Casaburi R, Barreiro E, Burelle Y, Debigaŕe R, et al. An official American thoracic society/european respiratory society statement: Update on limb muscle dysfunction in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2014;189(9):15–62.

258. Young P, Dewse M, Fergusson W, Kolbe J. Improvements in outcomes for chronic obstructive pulmonary disease (COPD) attributable to a hospital-based respiratory rehabilitation programme. [Internet]. Vol. 29, Australian and New Zealand journal of medicine. 1999.

259. Marciniuk DD, Brooks D, Butcher S, Debigare R, Dechman G, Ford G, et al. Optimizing pulmonary rehabilitation in chronic obstructive pulmonary disease – practical issues: A Canadian Thoracic Society Clinical Practice Guideline. Can Respir J. 2010;17(4).

260. Griffiths TL, Burr ML, Campbell IA, Lewis-Jenkins V, Mullins J, Shiels K, et al. Results at 1 year of outpatient multidisciplinary pulmonary rehabilitation: a randomised controlled trial. [Internet]. Vol. 355, Lancet. 2000.

261. Griffiths TL, Phillips CJ, Davies S, Burr ML, Campbell I a. Cost effectiveness of an outpatient multidisciplinary pulmonary rehabilitation programme. Thorax. 2001;56(10):779–84.

262. Global Initiative for Chronic Obstructive Lung Disease. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. 2017.

263. Medical Research Council Working Party. Long term domiciliary oxygen therapy in chronic hypoxic cor pulmonale complicating chronic bronchitis and emphysema. Lancet. 1981;317(8222):681–6.

264. Croxton TL, Bailey WC. Long-term oxygen treatment in chronic obstructive pulmonary disease: recommendations for future research: An NHLBI workshop report. Am J Respir Crit Care Med. 2006;174(4):373–8.

265. The Long-Term Oxygen Treatment Trial Research Group. A Randomized Trial of Long-Term Oxygen for COPD with Moderate Desaturation. N Engl J Med. 2016;375(17):1617–27.

266. Mehran RJ, Deslauriers J. Indications for surgery and patient work-up for bullectomy. Chest Surg Clin N Am. 1995;5(4):717–34.

267. Naunheim KS, Wood DE, Mohsenifar Z, Sternberg AL, Criner GJ, DeCamp MM, et al. Long-term follow-up of patients receiving lung-volume-reduction surgery versus medical therapy for severe emphysema by the National Emphysema Treatment Trial Research Group. Ann Thorac Surg. 2006;82(2):431–43.

268. the alpha-1-antitryspin deficiency registry study group. Survival and FEV 1 Decline in Individuals with Severe Deficiency of alpha 1 -Antitrypsin. Crit Care Med. 1998;158:49–59.

269. Seersholm N, Wencker M, Banik N, Viskum K, Dirksen A, Kok-Jensen A, et al. Does α1-antitrypsin augmentation therapy slow the annual decline in FEV1 in patients with severe hereditary α1-antitrypsin deficiency? Eur Respir J. 1997;10(10):2260–3.

270. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004;431(7011):931–45.

271. Ezkurdia I, Juan D, Rodriguez JM, Frankish A, Diekhans M, Harrow J, et al. Multiple evidence strands suggest that theremay be as few as 19 000 human protein-coding genes. Hum Mol Genet. 2014;23(22):5866–78.

272. Pertea M, Salzberg SL. Between a chicken and a grape: estimating the number of human genes. Genome Biol. 2010;11(5):206.

273. Harrow J, Nagy A, Reymond A, Alioto T, Patthy L, Antonarakis SE, et al. Identifying protein-coding genes in genomic sequences. Genome Biol. 2009;10(1):201.

274. D'Onofrio DJ, Abel DL. Redundancy of the genetic code enables translational pausing. Front Genet. 2014;5(MAY):1–16.

275. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, et al. The diploid genome sequence of an individual human. PLoS Biol. 2007;5(10):2113–44.

276. Collins F, Collins F, Brooks L, Brooks L, Chakravarti a, Chakravarti a. A DNA Polymorphism Discovery Resource for Research on Human Genetic Variation. Genome Res. 1998;8(12):1229–31.

277. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68–74.

278. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, et al. Mapping copy number variation by population-scale genome sequencing. Nature. 2011;470(7332):59–65.

279. Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. Nat Rev Genet. 2015;16(3):172–83.

280. Heller MJ. DNA Microarray Technology: Devices, Systems, and Applications. Annu Rev Biomed Eng. 2002;4(1):129–53.

281. Trevino V, Falciani F, Barrera-Saldana HA. DNA microarrays: a powerful genomic tool for biomedical and clinical research. Mol Med. 2007;13(9–10):527–41.

282. Bush WS, Moore JH. Chapter 11: Genome-Wide Association Studies. PLoS Comput Biol. 2012;8(12).

283. Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. Methods of integrating data to uncover genotype – phenotype interactions. Nat Publ Gr. 2015;(January).

284. Sanger F, Nicklen S, Coulson a R. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74(12):5463–7.

285. Manolio T a. Genomewide association studies and assessment of the risk of disease. N Engl J Med. 2010;363:166–76.

286. Huang T, Shu Y, Cai Y. Genetic differences among ethnic groups. BMC Genomics. 2015;16(1):1093.

287. Li M, Li C, Guan W. Evaluation of coverage variation of SNP chips for genome-wide association studies. Eur J Hum Genet. 2008;16(5):635–43.

288. Wagner MJ. Rare-variant genome-wide association studies: a new frontier in genetic ana lysis of complex traits. Pharmacogenomics. 2013;14:413–24.

289. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science (80-). 2012;337(6099):1190–5.

290. Ward LD, Kellis M. Interpreting noncoding genetic variation in complex traits and human disease. Nat Biotechnol. 2012;30(11):1095–106.

291. Zhang F, Lupski JR. Non-coding genetic variants in human disease. Hum Mol Genet. 2015;24(R1):R102–10.

292. Wen X, Luca F, Pique-Regi R. Cross-Population Joint Analysis of eQTLs: Fine Mapping and Functional Annotation. PLOS Genet. 2015;11(4):e1005176.

293. Kapoor A, Sekar RB, Hansen NF, Fox-Talbot K, Morley M, Pihur V, et al. An enhancer polymorphism at the cardiomyocyte intercalated disc protein NOS1AP locus is a major regulator of the QT interval. Am J Hum Genet. 2014;94(6):854–69.

294. Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, et al. Mapping the Genetic Architecture of Gene Expression in Human Liver. PLoS Biol. 2008;6(5):13.

295. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. Nat Publ Gr. 2015;16(February):197–212.

296. Lee MNHMN, Ye C, Villani a.-CA-C, Raj T, Li W, Eisenhaure TM, et al. Common Genetic Variants Modulate Pathogen-Sensing Responses in Human Dendritic Cells. Science (80-). 2014;343(6175):1246980.

297. Grundberg E, Small KS, Hedman AK, Nica AC, Buil A, Keildson S, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat Genet. 2012;44(10):1084–9.

298. Ding J, Gudjonsson JE, Liang L, Stuart PE, Li Y, Chen W, et al. Gene expression in skin and lymphoblastoid cells: Refined statistical method reveals extensive overlap in cis-eQTL signals. Am J Hum Genet. 2010;87(6):779–89.

299. Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. Science. 2009;325(5945):1246–50.

300. Fu J, Wolfs MGM, Deelen P, Westra H-J, Fehrmann RSN, Te Meerman GJ, et al. Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. PLoS Genet. 2012;8(1):e1002431.

301. Battle A, Khan Z, Wang SH, Mitrano A, Ford MJ, Pritchard JK, et al. Impact of regulatory variation from RNA to protein. Science (80-). 2014;347(6222):664–7.

302. Marsh SE, Travers J, Weatherall M, Williams M V, Aldington S, Shirtcliffe PM, et al. Proportional classifications of COPD phenotypes. Thorax. 2008;63(9):761–7.

303. Smith C a D, Harrison DJ. Association between polymorphism in gene for microsomal epoxide hydrolase and susceptibility to emphysema. Lancet. 1997;350(9078):630–3.

304. Laurell C-B, Eriksson S. The Electrophoretic α; 1 -Globulin Pattern of Serum in α; 1 -Antitrypsin Deficiency. Scand J Clin Lab Invest. 1963;15(2):132–40.

305. Greene CM, Marciniak SJ, Teckman J, Ferrarotti I, Brantly ML, Lomas DA, et al. α1-Antitrypsin deficiency. Nat Rev Dis Prim. 2016;2(July):16051.

306. Padrines M, Schneider-Pozzer M, Bieth JG. Inhibition of neutrophil elastase by alpha-1-proteinase inhibitor oxidized by activated neutrophils. Am Rev Respir Dis.

1989;139(3):783–90.

307. Needham M. Alpha-1-Antitrypsin deficiency * 3: Clinical manifestations and natural history. Thorax. 2004;59(5):441–5.

308. Eden E, Mitchell D, Mehlman B, Khouli H, Nejat M, Grieco MH, et al. Atopy, asthma, and emphysema in patients with severe ??-1-antitrypysin deficiency. Am J Respir Crit Care Med. 1997;156(1):68–74.

309. Holdt LM, Teupser D, Huber RM. Identification of a novel SERPINA-1 mutation causing alpha-1 antitrypsin deficiency in a patient with severe bronchiectasis and pulmonary embolism. Int J COPD. 2015;10:891–7.

310. Sveger T, Eriksson S. The liver in adolescents with alpha 1-antitrypsin deficiency. Hepatology. 1995;22(2):514–7.

311. Sveger T. The natural history of liver disease in alpha 1-antitrypsin deficient children. Acta Paediatr Scand. 1988;77(6):847–51.

312. Luisetti M, Seersholm N. Alpha1-antitrypsin deficiency. 1: epidemiology of alpha1-antitrypsin deficiency. Thorax. 2004;59(2):164–9.

313. Stockley R. Alpha1-Antitrypsin: more than just deficiency. Thorax. 2004;59(5):363–4.

314. DeMeo DL, Silverman EK. Alpha1-antitrypsin deficiency. 2: genetic aspects of alpha(1)-antitrypsin deficiency: phenotypes and genetic modifiers of emphysema risk. Thorax. 2004;59:259–64.

315. Hersh CP, Dahl M, Ly NP, Berkey CS, Nordestgaard BG, Silverman EK. Chronic obstructive pulmonary disease in alpha1-antitrypsin PI MZ heterozygotes: a meta-analysis. Thorax. 2004;59(10):843–9.

316. Bashir A, Shah NN, Hazari YM, Habib M, Bashir S, Hilal N, et al. Novel variants of SERPIN1A gene: Interplay between alpha1-antitrypsin deficiency and chronic obstructive pulmonary disease. Respir Med. 2016;117:139–49.

317. de Seynes C, Ged C, de Verneuil H, Chollet N, Balduyck M, Raherison C. Identification of a novel alpha1-antitrypsin variant. Respir Med Case Reports. 2017;20:64–7.

318. Graham RP, Dina MA, Howe SC, Butz ML, Willkomm KS, Murray DL, et al. SERPINA1 Full-Gene Sequencing Identifies Rare Mutations Not Detected in Targeted Mutation Analysis. J Mol Diagnostics. 2015;17(6):689–94.

319. Silverman GA, Pak SC, Perlmutter DH. Disorders of Protein Misfolding: Alpha-1-antitrypsin Deficiency as Prototype. J Pediatr. 2013;163(2):320–6.

320. Lace B, Sveger T, Krams A, Cernevska G, Krumina A. Age of SERPINA1 gene PI Z mutation: Swedish and Latvian population analysis. Ann Hum Genet. 2008;72(Pt 3):300–4.

321. Cox DW, Woo SLC, Mansfield T. DNA restriction fragments associated with alpha1-antitrypsin indicate a single origin for deficiency allele PI Z. Nature. 1985;316:79–81.

322. Seixas S, Garcia O, Trovoada JM, Santos TM, Amorim A, Rocha J. Patterns of haplotype diversity within the serpin gene cluster at 14q32.1: Insights into the natural history of the alpha1-antitrypsin polymorphism. Hum Genet. 2001;108(1):20–30.

323. Byth BC, Billingsley GD, Cox DW. Physical and genetic mapping of the serpin gene cluster at 14q32.1: allelic association and a unique haplotype associated with alpha 1-antitrypsin deficiency. Am J Hum Genet. 1994;55(1):126–33.

324. de Serres FJ, Blanco I, Fernández-Bustillo E. PI S and PI Z alpha-1 antitrypsin deficiency worldwide. A review of existing genetic epidemiological data. [Internet]. Vol. 67, Monaldi archives for chest disease = Archivio Monaldi per le malattie del torace / Fondazione clinica del lavoro, IRCCS [and] Istituto di clinica tisiologica e malattie apparato respiratorio, Università di Napoli, Secondo ateneo. 2007. 184-208 p.

325. Bossé Y. Genetics of chronic obstructive pulmonary disease: a succinct review, future avenues and prospective clinical applications. Pharmacogenomics. 2009;10(4):655–67.

326. Patnala R, Clements J, Batra J, Kwon J, Goate A, Collins F, et al. Candidate gene association studies: a comprehensive guide to useful in silico tools. BMC Genet. 2013;14(1):39.

327. Chen W, Brehm JM, Manichaikul A, Cho MH, Boutaoui N, Yan Q, et al. A Genome-Wide Association Study of Chronic Obstructive Pulmonary Disease in Hispanics. Ann Am Thorac Soc. 2015;12(3):340–8.

328. Cho M, McDonald M, Zhou X. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. Lancet Respir …. 2014;

329. Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP, et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. Nat Genet. 2010;42(3):200–2.

330. Cho MH, Castaldi PJ, Wan ES, Siedlinski M, Hersh CP, Demeo DL, et al. A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. Hum Mol Genet. 2012 Feb 15;21(4):947–57.

331. Hobbs BD, Parker MM, Chen H, Lao T, Hardin M, Qiao D, et al. Exome Array

Analysis Identifies A Common Variant in IL27 Associated with Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2016;194(1):48–57.

332. Hobbs BD, Jong K de, Lamontagne M, Bossé Y, Shrine N, Artigas MS, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. Nat Genet. 2017;49(3):426–32.

333. Wilk JB, Walter RE, Laramie JM, Gottlieb DJ, O'Connor GT. Framingham Heart Study genome-wide association: results for pulmonary function measures. BMC Med Genet. 2007;8(Suppl 1):S8.

334. Wilk JB, Chen T, Gottlieb DJ, Walter RE, Nagle MW, Brandler BJ, et al. A Genome-Wide Association Study of Pulmonary Function Measures in the Framingham Heart Study. PLoS Genet. 2009;5(3):8.

335. Wan ES, Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, et al. Genome-wide association analysis of body mass in chronic obstructive pulmonary disease. Am J Respir Cell Mol Biol. 2011 Aug;45(2):304–10.

336. Wain L V, Shrine N, Miller S, Jackson VE, Ntalla I, Artigas MS, et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. Lancet Respir Med. 2015;i(15):1–13.

337. van der Plaat DA, de Jong K, Lahousse L, Faiz A, Vonk JM, van Diemen CC, et al. Genome-wide association study on the FEV1/FVC-ratio in never-smokers identifies HHIP and FAM13A. J Allergy Clin Immunol. 2016;1–8.

338. Thyagarajan B, Wojczynski M, Minster RL, Sanders J, Barral S, Christiansen L, et al. Genetic variants associated with lung function: the long life family study. Respir Res. 2014;15:134.

339. Soler Artigas M, Loth D, Wain L, Gharib S, Obeidat M, Tang W, et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. Nat Genet. 2011;43(11):1082–90.

340. Repapi E, Sayers I, Wain L V, Burton PR, Johnson T, Obeidat M, et al. Genome-wide association study identifies five loci associated with lung function. Nat Genet. 2010;42(1):36–44.

341. Manichaikul A, Hoffman E a, Smolonska J, Gao W, Cho MH, Baumhauer H, et al. Genome-wide study of percent emphysema on computed tomography in the general population. The Multi-Ethnic Study of Atherosclerosis Lung/SNP Health Association Resource Study. Am J Respir Crit Care Med. 2014;189(4):408–18.

342. Lutz SM, Cho MH, Young K, Hersh CP, Castaldi PJ, McDonald M-L, 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.

343. Loth DW, Artigas MS, Gharib S a, Wain L V, Franceschini N, Koch B, et al. Genome-wide association analysis identifies six new loci associated with forced vital capacity. Nat Genet. 2014;46(7):669–77.

344. Lee J, Cho MH, Hersh CP, McDonald M-LN, Crapo JD, Bakke PS, et al. Genetic susceptibility for chronic bronchitis in chronic obstructive pulmonary disease. Respir Res. 2014;15(1):113.

345. Kong X, Cho MH, Anderson W, Coxson HO, Muller N, Washko G, et al. Genome-wide Association Study Identifies BICD1 as a Susceptibility Gene for Emphysema. [Internet]. Vol. 183, American Journal of Respiratory and Critical Care Medicine. 2011.

346. Kim WJ, Lee MK, Shin C, Cho NH, Lee S Do, Oh Y-MM, et al. Genome-wide association studies identify locus on 6p21 influencing lung function in the Korean population. Respirology. 2014 Jan 6;19(3):1–9.

347. Kim DK, Cho MH, Hersh CP, Lomas D a, Miller BE, Kong X, et al. Genome-Wide Association Analysis of Blood Biomarkers in Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2012 Dec 15;186(12):1238–47.

348. Jong K de, Vonk JM, Faiz A, Plaat DA van der, Timens W, Bossé Y, et al. Novel Genetic Susceptibility Loci for FEV1 in the Context of Occupational Exposure in Never-Smokers. Am J Respir Crit Care Med. 2016;194(6):769–72.

349. Imboden M, Bouzigon E, Curjuric I, Ramasamy A, Kumar A, Hancock DB, et al. Genome-wide association study of lung function decline in adults with and without asthma. J Allergy Clin Immunol. 2012 May;129(5):1–11.

350. Hansel NN, Ruczinski I, Rafaels N, Sin DD, Daley D, Malinina A, et al. Genome-wide study identifies two loci associated with lung function decline in mild to moderate COPD. Hum Genet. 2013 Sep 18;132(1):79–90.

351. Hansel NN, Paré PD, Rafaels N, Sin D, Sandford A, Daley D, et al. Genome Wide Association Study Identifies Novel Loci Associated with Airway Responsiveness in COPD. Am J Respir Cell Mol Biol. 2014;53(2):226–34.

352. Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marciante KD, et al. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. Nat Genet. 2010;42(1):45–52.

353. Hancock DB, Artigas MS, Gharib S a., Henry A, Manichaikul A, Ramasamy A, et al. Genome-Wide Joint Meta-Analysis of SNP and SNP-by-Smoking Interaction Identifies Novel Loci for Pulmonary Function. PLoS Genet. 2012;8(12):e1003098.

354. Dijkstra AE, Smolonska J, Van Den Berge M, Wijmenga C, Zanen P, Luinge MA, et al. Susceptibility to chronic mucus hypersecretion, a genome wide association study. PLoS One. 2014;9(4).

355. Dijkstra AE, Postma DS, van Ginneken B, Wielpütz MO, Schmidt M, Becker N, et al. Novel genes for airway wall thickness identified with combined genome-wide association and expression analyses. Am J Respir Crit Care Med. 2015;191(5):547–56.

356. De Jong K, Vonk JM, Timens W, Bossé Y, Sin DD, Hao K, et al. Genome-wide interaction study of gene-by-occupational exposure and effects on FEV1 levels. J Allergy Clin Immunol. 2015;136(6):1664–1672e14.

357. Cho MH, Castaldi PJ, Hersh CP, Hobbs BD, Barr RG, Tal-Singer R, et al. A Genome-Wide Association Study of Emphysema and Airway Quantitative Imaging Phenotypes. Am J Respir Crit Care Med. 2015;192(5):559–69.

358. Castaldi PJ, Cho MH, Est??par RSJ, McDonald MLN, Laird N, Beaty TH, et al. Genome-wide association identifies regulatory loci associated with distinct local histogram emphysema patterns. Am J Respir Crit Care Med. 2014;190(4):399–409.

359. Artigas MS, Wain L V., Miller S, Kheirallah AK, Huffman JE, Ntalla I, et al. Sixteen new lung function signals identified through 1000 Genomes Project reference panel imputation. Nat Commun. 2015;6:8658.

360. Wain L V, Shrine N, Artigas MS, Erzurumluoglu AM, Noyvert B, Bossini-Castillo L, et al. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. Nat Genet. 2017;49(3):416–25.

361. Lee JH, Cho MH, Hersh CP, McDonald M-LN, Wells JM, Dransfield MT, et al. IREB2 and GALC are associated with pulmonary artery enlargement in chronic obstructive pulmonary disease. Am J Respir Cell Mol Biol. 2015;52(3):365–76.

362. Corvol H, Hodges CA, Drumm ML, Guillot L. Moving beyond genetics: is FAM13A a major biological contributor in lung physiology and chronic lung diseases? J Med Genet. 2014;51(10):646–9.

363. Lamontagne M, Couture C, Postma DS, Timens W, Sin DD, Paré PD, et al. Refining Susceptibility Loci of Chronic Obstructive Pulmonary Disease with Lung eqtls. PLoS One. 2013 Jul 30;8(7):e70220.

364. Jiang Z, Lao T, Qiu W, Polverino F, Gupta K, Guo F, et al. A Chronic Obstructive Pulmonary Disease Susceptibility Gene, FAM13A, Regulates Protein Stability of β-Catenin. Am J Respir Crit Care Med. 2016;194(2):185–97.

365. Zhou X, Baron RM, Hardin M, Cho MH, Zielinski J, Hawrylkiewicz I, et al. Identification of a chronic obstructive pulmonary disease genetic determinant that regulates HHIP. Hum Mol Genet. 2012;21(6):1325–35.

366. Zhou X, Qiu W, Sathirapongsasuti JF, Cho MH, Mancini JD, Lao T, et al. Gene expression analysis uncovers novel Hedgehog interacting protein (HHIP) effects in

human bronchial epithelial cells. Genomics. 2013 Feb 28;101(5):263–72.

367. Annoni R, Lanças T, Yukimatsu Tanigawa R, de Medeiros Matsushita M, de Morais Fernezlian S, Bruno A, et al. Extracellular matrix composition in COPD. Eur Respir J. 2012;40(6):1362–73.

368. Sorci G, Riuzzi F, Giambanco I, Donato R. RAGE in tissue homeostasis, repair and regeneration. Biochim Biophys Acta - Mol Cell Res. 2013;1833(1):101–9.

369. Wu L, Ma L, Nicholson LFB, Black PN. Advanced glycation end products and its receptor (RAGE) are increased in patients with COPD. Respir Med. 2011;105(3):329–36.

370. Kim WJ, Lee S Do. Candidate genes for COPD: current evidence and research. Int J Chron Obstruct Pulmon Dis. 2015;10:2249–55.

371. David SP, Hamidovic A, Chen GK, Bergen a W, Wessel J, Kasberger JL, et al. Genome-wide meta-analyses of smoking behaviors in African Americans. Transl Psychiatry. 2012 Jan;2(March):e119.

372. Furberg H, Kim Y, Dackor J, Boerwinkle E, Franceschini N, Ardissino D, et al. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat Genet. 2010;42(5):441–7.

373. Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton L, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. Nat Genet. 2010;42(5):436–40.

374. Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, et al. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. Nat Genet. 2010;42(5):448–53.

375. Broderick P, Wang Y, Vijayakrishnan J, Matakidou A, Spitz MR, Eisen T, et al. Deciphering the impact of common genetic variation on lung cancer risk: A genome-wide association study. Cancer Res. 2009;69(16):6633–41.

376. Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, et al. A Genome-wide Association Study of Lung Cancer Identifies a Region of Chromosome 5p15 Associated with Risk for Adenocarcinoma. Am J Hum Genet. 2009;85(5):679–91.

377. Wang Y, Broderick P, Webb E, Wu X, Vijayakrishnan J, Matakidou A, et al. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. Nat Genet. 2008;40(12):1407–9.

378. Liu P, Vikis HG, Wang D, Lu Y, Wang Y, Schwartz AG, et al. Familial aggregation of common sequence variants on 15q24-25.1 in lung cancer. J Natl Cancer Inst. 2008;100(18):1326–30.

379. McKay JD, Hung RJ, Gaborieau V, Boffetta P, Chabrier A, Byrnes G, et al. Lung cancer susceptibility locus at 5p15.33. Nat Genet. 2008;40(12):1404–6.

380. Brunzell DH, Picciotto MR. Molecular mechanisms underlying the motivational effects of nicotine. Nebr Symp Motiv. 2009;55:17–30.

381. Steinlein OK, Bertrand D. Neuronal nicotinic acetylcholine receptors: From the genetic analysis to neurological diseases. Biochem Pharmacol. 2008;76(10):1175–83.

382. Arja C, Ravuri RR, Pulamaghatta VN, Surapaneni KM, Raya P, Adimoolam C, et al. Genetic determinants of chronic obstructive pulmonary disease in south Indian male smokers. PLoS One. 2014;9(2):1–6.

383. DeMeo DL, Mariani T, Bhattacharya S, Srisuma S, Lange C, Litonjua A, et al. Integration of Genomic and Genetic Approaches Implicates IREB2 as a COPD Susceptibility Gene. Am J Hum Genet. 2009;85(4):493–502.

384. Du YM, Xue YW, Xiao W. Association of IREB2 Gene rs2568494 Polymorphism with Risk of Chronic Obstructive Pulmonary Disease: A Meta-Analysis. Med Sci Monit. 2016;22:177–82.

385. Wilk JB, Djousse L, Arnett DK, Rich SS, Province MA, Hunt SC, et al. Evidence for major genes influencing pulmonary function in the NHLBI Family Heart Study. Genet Epidemiol. 2000;19(1):81–94.

386. Palmer LJ, Knuiman MW, Divitini ML, Burton PR, James AL, Bartholomew HC, et al. Familial aggregation and heritability of adult lung function: Results from the Busselton health study. Eur Respir J. 2001;17(4):696–702.

387. Qiu W, Cho MH, Riley JH, Anderson WH, Singh D, Bakke P, et al. Genetics of sputum gene expression in chronic obstructive pulmonary disease. PLoS One. 2011 Jan;6(9):e24395.

388. Castaldi PJP, Cho MH, Zhou X, Qiu W, Mcgeachie M, Celli B, et al. Genetic control of gene expression at novel and established chronic obstructive pulmonary disease loci. Hum Mol Genet. 2014;24(4):1200–10.

389. Lamontagne M, Joubert P, Timens W, Postma DS, Hao K, Nickle D, et al. Susceptibility genes for lung diseases in the major histocompatibility complex revealed by lung expression quantitative trait loci analysis. Eur Respir J. 2016;48(2):573–6.

390. Matsson H, Söderhäll C, Einarsdottir E, Lamontagne M, Gudmundsson S, Backman H, et al. Targeted high-throughput sequencing of candidate genes for chronic obstructive pulmonary disease. BMC Pulm Med. 2016;16(1):146.

391. Hao K, Bossé Y, Nickle DC, Paré PD, Postma DS, Laviolette M, et al. Lung eQTLs

to Help Reveal the Molecular Underpinnings of Asthma. PLoS Genet. 2012;8(11):e1003029.

392. Nguyen JDU, Lamontagne M, Couture C, Conti M, Paré PD, Sin DD, et al. Susceptibility loci for lung cancer are associated with mRNA levels of nearby genes in the lung. Carcinogenesis. 2014;35(12):2653–9.

393. Saferali A, Obeidat M, Bérubé J-C, Lamontagne M, Bossé Y, Laviolette M, et al. Polymorphisms Associated with Expression of BPIFA1/BPIFB1 and Lung Disease Severity in Cystic Fibrosis. Am J Respir Cell Mol Biol. 2015;53(5):607–14.

394. Bérubé J, Lamontagne M, Couture C, Nickle D, Timens W, Postma D, et al. Genome-wide genetic ancestry measurements to predict lung function in European populations. Eur Respir J. 2013 Oct;42(4):1144–7.

395. Lamontagne M, Timens W, Hao K, Bossé Y, Laviolette M, Steiling K, et al. Genetic regulation of gene expression in the lung identifies CST3 and CD22 as potential causal genes for airflow obstruction. Thorax. 2014;69(11):997–1004.

396. Obeidat M, Miller S, Probert K, Billington CK, Henry AP, Hodge E, et al. GSTCD and INTS12 Regulation and Expression in the Human Lung. PLoS One. 2013 Jan;8(9):e74630.

397. Altshuler D, Daly MJ, Lander E. Genetic Mapping in Human Disease. Science (80-). 2009;322(5903):881–8.

398. Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK, et al. Gene map of the extended human MHC. Nat Rev Genet. 2004;5(12):889–99.

399. de Bakker PIW, Raychaudhuri S. Interrogating the major histocompatibility complex with high-throughput genomics. Hum Mol Genet. 2012;21(R1):29–36.

400. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med. 2010;363(13):1211–21.

401. Noguchi E, Sakamoto H, Hirota T, Ochiai K, Imoto Y, Sakashita M, et al. Genome-wide association study identifies HLA-DP as a susceptibility gene for pediatric asthma in Asian populations. PLoS Genet. 2011;7(7):e1002170.

402. Ramasamy A, Kuokkanen M, Vedantam S, Gajdos ZK, Couto Alves A, Lyon HN, et al. Genome-Wide Association Studies of Asthma in Population-Based Cohorts Confirm Known and Suggested Loci and Identify an Additional Association near HLA. PLoS One. 2012 Jan;7(9):e44008.

403. Lasky-Su J, Himes BE, Raby B a, Klanderman BJ, Sylvia JS, Lange C, et al. HLA-DQ strikes again: Genome-wide association study further confirms HLA-DQ in the diagnosis of asthma among adults. Clin Exp Allergy. 2012;42(12):1724–33.

404. Ferreira M a R, Matheson MC, Tang CS, Granell R, Ang W, Hui J, et al. Genome-wide association analysis identifies 11 risk variants associated with the asthma with hay fever phenotype. J Allergy Clin Immunol. 2014;133(6):1564–71.

405. Hirota T, Takahashi A, Kubo M, Tsunoda T, Tomita K, Doi S, et al. Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. Nat Genet. 2011;43(9):893–6.

406. Lan Q, Hsiung C a, Matsuo K, Hong Y-C, Seow A, Wang Z, et al. Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. Nat Genet. 2012 Nov 11;44(12):1330–5.

407. Shiraishi K, Kunitoh H, Daigo Y, Takahashi A, Goto K, Sakamoto H, et al. A genome-wide association study identifies two new susceptibility loci for lung adenocarcinoma in the Japanese population. Nat Genet. 2012 Jul 15;44(8):900–3.

408. Wright F a, Strug LJ, Doshi VK, Commander CW, Blackman SM, Sun L, et al. Genome-wide association and linkage identify modifier loci of lung disease severity in cystic fibrosis at 11p13 and 20q13.2. Nat Genet. 2011;43(6):539–46.

409. Fingerlin TE, Murphy E, Zhang W, Peljto a L, Brown KK, Steele MP, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. Nat Genet. 2013;45(6):613–20.

410. Vassallo R, Walters PR, Lamont J, Kottom TJ, Yi ES, Limper AH. Cigarette smoke promotes dendritic cell accumulation in COPD; a Lung Tissue Research Consortium study. Respir Res. 2010;11:45.

411. Zhang J, He J, Xia J, Chen Z, Chen X. Delayed apoptosis by neutrophils from COPD patients is associated with altered bak, bcl-xl, and mcl-1 mRNA expression. Diagn Pathol. 2012;7(1):65.

412. Davies G, Marioni RE, Liewald DC, Hill WD, Hagenaars SP, Harris SE, et al. Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N= 112 151). Mol Psychiatry. 2016;21(August 2015):1–10.

413. Ibrahim-Verbaas C a, Bressler J, Debette S, Schuur M, Smith a V, Bis JC, et al. GWAS for executive function and processing speed suggests involvement of the CADM2 gene. Mol Psychiatry. 2015;1–9.

414. Stringer S, Minică CC, Verweij KJH, Mbarek H, Bernard M, Derringer J, et al. Genome-wide association study of lifetime cannabis use based on a large meta-analytic sample of 32 330 subjects from the International Cannabis Consortium. Transl Psychiatry. 2016;6(December 2015):e769.

415. Butler MW, Fukui T, Salit J, Shaykhiev R, Mezey JG, Hackett NR, et al. Modulation of cystatin A expression in human airway epithelium related to genotype, smoking, COPD, and lung cancer. Cancer Res. 2011 Apr 1;71(7):2572–81.

416. Blaydon DC, Nitoiu D, Eckl KM, Cabral RM, Bland P, Hausser I, et al. Mutations in CSTA, encoding cystatin A, underlie exfoliative ichthyosis and reveal a role for this protease inhibitor in cell-cell adhesion. Am J Hum Genet. 2011;89(4):564–71.

417. Ware LB, Matthay MA. Keratinocyte and hepatocyte growth factors in the lung: roles in lung development, inflammation, and repair. Am J Physiol Lung Cell Mol Physiol. 2002;282(5):L924-40.

418. Dial CF, Tune MK, Doerschuk CM, Mock JR. Foxp3+ Regulatory T Cell Expression of Keratinocyte Growth Factor Enhances Lung Epithelial Proliferation. Am J Respir Cell Mol Biol. 2017;

419. Sauleda J, Noguera A, Blanquer D, Pons J, López M, Villena C, et al. Pulmonary and systemic hepatocyte and keratinocyte growth factors in patients with chronic obstructive pulmonary disease. Int J Chron Obstruct Pulmon Dis. 2008;3(4):719–25.

420. Hao S, Long F, Sun F, Liu T, Li D, Jiang S. Birt–Hogg–Dubé syndrome: a literature review and case study of a Chinese woman presenting a novel FLCN mutation. BMC Pulm Med. 2017;17(1):43.

421. Painter JN, Tapanainen H, Somer M, Tukiainen P, Aittomäki K. A 4-bp deletion in the Birt-Hogg-Dubé gene (FLCN) causes dominantly inherited spontaneous pneumothorax. Am J Hum Genet. 2005;76(3):522–7.

422. Cho MH, Klanderman BJ, Litonjua A a, Sparrow D, Silverman EK, Raby B a. Folliculin mutations are not associated with severe COPD. BMC Med Genet. 2008;9:120.

423. Nakajima J, Takamoto S, Murakawa T, Fukami T, Yoshida Y, Kusakabe M. Outcomes of thoracoscopic management of secondary pneumothorax in patients with COPD and interstitial pulmonary fibrosis. Surg Endosc Other Interv Tech. 2009;23(7):1536–40.

424. Collins SA, Lucas JSA, Inskip HM, Godfrey KM, Roberts G, Holloway JW. HHIP, HDAC4, NCR3 and RARB polymorphisms affect fetal, childhood and adult lung function. Eur Respir J. 2013;41(3):756–7.

425. Fend L, Rusakiewicz S, Adam J, Bastien B, Caignard A, Messaoudene M, et al. Prognostic impact of the expression of NCR1 and NCR3 NK cell receptors and PD-L1 on advanced non-small cell lung cancer. Oncoimmunology. 2017;6(1):e1163456.

426. Truneh A, Sharma S, Silverman C, Khandekar S, Reddy MP, Deen KC, et al. Temperature-sensitive differential affinity of TRAIL for its receptors: DR5 is the highest affinity receptor. J Biol Chem. 2000;275(30):23319–25.

427. Schneider P, Thome M, Burns K, Bodmer JL, Hofmann K, Kataoka T, et al. TRAIL receptors 1 (DR4) and 2 (DR5) signal FADD-dependent apoptosis and activate NF-κB. Immunity. 1997;7(6):831–6.

428. Kischkel FC, Lawrence DA, Chuntharapai A, Schow P, Kim KJ, Ashkenazi A. Apo2L/TRAIL-Dependent Recruitment of Endogenous FADD and Caspase-8 to Death Receptors 4 and 5. Immunity. 2000;12(6):611–20.

429. Solarewicz-Madejek K, Basinski TM, Crameri R, Akdis M, Akkaya A, Blaser K, et al. T cells and eosinophils in bronchial smooth muscle cell death in asthma. Clin Exp Allergy. 2009;39(6):845–55.

430. Morissette MC, Vachon-Beaudoin G, Parent J, Chakir J, Milot J. Increased p53 Level, Bax/Bcl-xL Ratio, and TRAIL Receptor Expression in Human Emphysema. Am J Respir Crit Care Med. 2008;178(3):240–7.

431. Wu Y, Shen Y, Zhang J, Wan C, Wang T, Xu D, et al. Increased serum TRAIL and DR5 levels correlated with lung function and inflammation in stable COPD patients. Int J Chron Obs Pulmon Dis. 2015;10:2405–12.

432. Land SC, Scott CL, Walker D. MTOR signalling, embryogenesis and the control of lung development. Semin Cell Dev Biol. 2014;36:68–78.

433. Nadon AM, Perez MJ, Hernandez-Saavedra D, Smith LP, Yang Y, Sanders LA, et al. Rtp801 suppression of epithelial mTORC1 augments endotoxin-induced lung inflammation. Am J Pathol. 2014;184(9):2382–9.

434. Khabibullin D, Medvetz D a, Pinilla M, Hariharan V, Li C, Hergrueter A, et al. Folliculin regulates cell-cell adhesion, AMPK, and mTORC1 in a cell-type-specific manner in lung-derived cells. Physiol Rep. 2014;2(8):1–12.

435. Galanter JM, Gignoux CR, Torgerson DG, Roth LA, Eng C, Oh SS, et al. Genome-wide association study and admixture mapping identify different asthma-associated loci in Latinos: The Genes-environments & Admixture in Latino Americans study. J Allergy Clin Immunol. 2014;134(2):295–305.

436. Hijikata M, Matsushita I, Tanaka G, Tsuchiya T, Ito H, Tokunaga K, et al. Molecular cloning of two novel mucin-like genes in the disease- susceptibility locus for diffuse panbronchiolitis. Hum Genet. 2011;129(2):117–28.

437. Baños-Lara MDR, Piao B, Guerrero-Plata A. Differential mucin expression by respiratory syncytial virus and human metapneumovirus infection in human epithelial cells. Mediators Inflamm. 2015;2015.

438. Furio L, Hovnanian A. Netherton syndrome: Defective kallikrein inhibition in the skin leads to skin inflammation and allergy. Biol Chem. 2014;395(9):945–58.

439. Hovnanian A. Netherton syndrome: Skin inflammation and allergy by loss of protease inhibition. Cell Tissue Res. 2013;351(2):289–300.

440. Birben E, Sackesen C, Turgutoğlu N, Kalayci Ö. The role of SPINK5 in asthma related physiological events in the airway epithelium. Respir Med. 2012;106(3):349–

55.

441. Demedts IK, Demoor T, Bracke KR, Joos GF, Brusselle GG. Role of apoptosis in the pathogenesis of COPD and pulmonary emphysema. Respir Res. 2006;7:53.

442. Gorer PA. The detection of a hereditary antigenic difference in the blood of mice by means of human group A serum. J Genet. 1936;32(1):17–31.

443. Bodd M, Kim C, Lundin KEA, Sollid LM. T-cell response to gluten in patients with HLA-DQ2.2 reveals requirement of peptide-MHC stability in celiac disease. Gastroenterology. 2012;142(3):552–61.

444. Allegretti YL, Bondar C, Guzman L, Cueto Rua E, Chopita N, Fuertes M, et al. Broad MICA/B Expression in the Small Bowel Mucosa: A Link between Cellular Stress and Celiac Disease. PLoS One. 2013;8(9).

445. Bär F, Sina C, Hundorfean G, Pagel R, Lehnert H, Fellermann K, et al. Inflammatory bowel diseases influence major histocompatibility complex class I (MHC I) and II compartments in intestinal epithelial cells. Clin Exp Immunol. 2013;172(2):280–9.

446. Goyette P, Boucher G, Mallon D, Ellinghaus E, Jostins L, Huang H, et al. High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. Nat Genet. 2015;47(2):172–9.

447. Sugiyama Y, Kudoh S, Maeda H, Suzaki H, Takaku F. Analysis of HLA antigens in patients with diffuse panbronchiolitis. Am Rev Respir Dis. 1990 Jun;141(6):1459–62.

448. Maranetra N, Chandanayingyong D, Bovornkitti S. HLA antigen and ventilatory drive in Thais with chronic obstructive pulmonary disease. Asian Pacific J Allergy Immunol. 1990;8(2):137–40.

449. Huang S, Su C, Chang S. Tumor necrosis factor-α gene polymorphism in chronic bronchitis. Am J Respir Crit Care Med. 1997;156:1436–9.

450. Keicho N, Tokunaga K, Nakata K, Taguchi Y, Azuma A, Tanabe K, et al. Contribution of TAP genes to genetic predisposition for diffuse panbronchiolitis. Am J Respir Crit Care Med. 1998 Apr;158(3):846–50.

451. Keatings VM, Cave SJ, Henry MJ, Morgan K, O'Connor CM, FitzGerald MX, et al. A polymorphism in the tumor necrosis factor-alpha gene promoter region may predispose to a poor prognosis in COPD. Chest. 2000;118(4):971–5.

452. Sakao S, Tatsumi K, Igari H, Shino Y, Shirasawa H, Kuriyama T. Association of tumor necrosis factor α gene promoter polymorphism with the presence of chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2001;163(2):420–2.

453. Kucukaycan M, Krugten M Van. Tumor necrosis factor-alpha+ 489G/A gene polymorphism is associated with chronic obstructive pulmonary disease. Respir Med. 2002;3(1):1–7.

454. Kulski JK, Dunn DS, Hui J, Martinez P, Romphruk A V, Leelayuwat C, et al. Alu polymorphism within the MICB gene and association with HLA-B alleles. Immunogenetics. 2002 Feb;53(10–11):975–9.

455. Ianbaeva DG, Korytina GF, Viktorova T V. Allelic variants of the tumor necrosis factor superfamily as markers of the severity of the course of chronic obstructive lung disease and bronchiectatic disease. Genetika. 2004;40(4):545–51.

456. Yang I a, Holz O, Jörres R a, Magnussen H, Barton SJ, Rodríguez S, et al. Association of tumor necrosis factor-alpha polymorphisms and ozone-induced change in lung function. Am J Respir Crit Care Med. 2005 Jan 15;171(2):171–6.

457. Randolph AG, Lange C, Silverman EK, Lazarus R, Weiss ST. Extended haplotype in the tumor necrosis factor gene cluster is associated with asthma and asthma-related phenotypes. Am J Respir Crit Care Med. 2005 Sep 15;172(6):687–92.

458. Jiang L, He B, Zhao M, Ning L, Li X, Yao W. Association of gene polymorphisms of tumour necrosis factor-alpha and interleukin-13 with chronic obstructive pulmonary disease in Han nationality in Beijing. Chin Med J (Engl). 2005;118(7):541–7.

459. Hersh CP, Demeo DL, Lange C, Litonjua A a, Reilly JJ, Kwiatkowski D, et al. Attempted replication of reported chronic obstructive pulmonary disease candidate gene associations. Am J Respir Cell Mol Biol. 2005 Jul;33(1):71–8.

460. Hui J, Palmer LJ, James a L, Musk a W, Beilby JP. AluyMICB dimorphism within the class I region of the major histocompatibility complex is associated with asthma and airflow obstruction in the Busselton population. Clin Exp Allergy. 2006 Jun;36(6):728–34.

461. Matheson MC, Ellis J a, Raven J, Walters EH, Abramson MJ. Association of IL8, CXCR2 and TNF-alpha polymorphisms and airway disease. J Hum Genet. 2006 Jan;51(3):196–203.

462. Ruse CE, Hill MC, Tobin M, Neale N, Connolly MJ, Parker SG, et al. Tumour necrosis factor gene complex polymorphisms in chronic obstructive pulmonary disease. Respir Med. 2007 Feb;101(2):340–4.

463. Zhang H, Hang J, Wang X, Zhou W, Sun B, Dai H, et al. TNF polymorphisms modify endotoxin exposure-associated longitudinal lung function decline. Occup Environ Med. 2007 Jun;64(6):409–13.

464. Curtis JL, Freeman CM, Hogg JC. The immunopathogenesis of chronic obstructive pulmonary disease: insights from recent research. Proc Am Thorac Soc. 2007 Oct

1;4(7):512–21.

465. KV Danilko, GF Korytina, LZ Akhmidishina, DG Ianbaeva, ShZ Zagidullin TV. Association of cytokines genes (ILL, IL1RN, TNF, LTA, IL6, IL8, IL0) polymorphic markers with chronic obstructive pulmonary disease. Mol Biol (Mosk). 2007;41(1):26–36.

466. Papatheodorou A, Latsi P, Vrettou C, Dimakou A, Chroneou A, Makrythanasis P, et al. Development of a novel microarray methodology for the study of SNPs in the promoter region of the TNF-alpha gene: their association with obstructive pulmonary disease in Greek patients. Clin Biochem. 2007 Aug;40(12):843–50.

467. Birrell MA, Wong S, Catley MC, Belvisi MG. Impact of tobacco-smoke on key signaling pathways in the innate immune response in lung macrophages. J Cell …. 2008;(December 2006):27–37.

468. Kim EY, Battaile JT, Patel AC, You Y, Agapov E, Grayson MH, et al. Persistent activation of an innate immune response translates respiratory viral infection into chronic lung disease. Nat Med. 2008 Jul;14(6):633–40.

469. Demeo DL, Campbell EJ, Barker AF, Brantly ML, Eden E, McElvaney NG, et al. IL10 polymorphisms are associated with airflow obstruction in severe alpha1-antitrypsin deficiency. Am J Respir Cell Mol Biol. 2008 Jan;38(1):114–20.

470. Gadgil A, Duncan SR. Role of T-lymphocytes and pro-inflammatory mediators in the pathogenesis of chronic obstructive pulmonary disease. Int J Chron Obstruct Pulmon Dis. 2008 Jan;3(4):531–41.

471. Gingo MR, Silveira LJ, Miller YE, Friedlander a L, Cosgrove GP, Chan ED, et al. Tumour necrosis factor gene polymorphisms are associated with COPD. Eur Respir J Off J Eur Soc Clin Respir Physiol. 2008 May;31(5):1005–12.

472. Yucesoy B, Kurzius-Spencer M, Johnson VJ, Fluharty K, Kashon ML, Guerra S, et al. Association of cytokine gene polymorphisms with rate of decline in lung function. J Occup Environ Med. 2008 Jun;50(6):642–8.

473. Stankovic MM, Nestorovic AR, Tomovic AM, Petrovic-Stanojevic ND, Andjelic-Jelic MS, Dopudja-Pantic VB, et al. TNF-alpha-308 promotor polymorphism in patients with chronic obstructive pulmonary disease and lung cancer. Neoplasma. 2009;56(4):348–52.

474. Koller B, Bals R, Roos D, Korting HC, Griese M, Hartl D. Innate immune receptors on neutrophils and their role in chronic lung disease. Eur J Clin Invest. 2009 Jul;39(7):535–47.

475. Andersson CK, Mori M, Bjermer L, Löfdahl C-G, Erjefält JS. Alterations in lung mast cell populations in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2010 Feb 1;181(3):206–17.

476. Polverino F, Baraldo S, Bazzan E, Agostini S, Turato G, Lunardi F, et al. A novel insight into adaptive immunity in chronic obstructive pulmonary disease: B cell activating factor belonging to the tumor necrosis factor family. Am J Respir Crit Care Med. 2010 Oct 15;182(8):1011–9.

477. Sapey E, Wood AM, Ahmad A, Stockley R a. Tumor necrosis factor-{alpha} rs361525 polymorphism is associated with increased local production and downstream inflammation in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2010 Jul 15;182(2):192–9.

478. Chen Y-C, Liu S-F, Chin C-H, Wu C-C, Chen C-J, Chang H-W, et al. Association of tumor necrosis factor-alpha-863C/A gene polymorphism with chronic obstructive pulmonary disease. Lung. 2010 Aug;188(4):339–47.

479. Castaldi PJ, Cho MH, Cohn M, Langerman F, Moran S, Tarragona N, et al. The COPD genetic association compendium: a comprehensive online database of COPD genetic associations. Hum Mol Genet. 2010;19(3):526–34.

480. Siedlinski M, Cho MH, Bakke P, Gulsvik A, Lomas D a, Anderson W, et al. Genome-wide association study of smoking behaviours in patients with COPD. Thorax. 2011 Oct;66(10):894–902.

481. Córdoba-Lanús E, Baz-Dávila R, de-Torres JP, Rodríguez-Pérez MC, Maca-Meyer N, Varo N, et al. TNFA-863 polymorphism is associated with a reduced risk of chronic obstructive pulmonary disease: a replication study. BMC Med Genet. 2011 Jan;12(1):132.

482. Pace E, Giarratano A, Ferraro M, Bruno A, Siena L, Mangione S, et al. TLR4 upregulation underpins airway neutrophilia in smokers with chronic obstructive pulmonary disease and acute respiratory failure. Hum Immunol. 2011 Jan;72(1):54–62.

483. Guo Y, Gong Y, Pan C, Qian Y, Shi G, Cheng Q, et al. Association of genetic polymorphisms with chronic obstructive pulmonary disease in the Chinese Han population: a case--control study. BMC Med Genomics. 2012 Dec 26;5(1):64.

484. Faner R, Nuñez B, Sauleda J, Garcia-Aymerich J, Pons J, Crespí C, et al. HLA Distribution in COPD Patients. COPD. 2013 Mar 20;138–46.

485. Yin X, Low HQ, Wang L, Li Y, Ellinghaus E, Han J, et al. Genome-wide meta-analysis identifies multiple novel associations and ethnic heterogeneity of psoriasis susceptibility. Nat Commun. 2015;6:6916.

486. Zhou F, Cao H, Zuo X, Zhang T, Zhang X, Liu X, et al. Deep sequencing of the MHC region in the Chinese population contributes to studies of complex disease. Nat Genet. 2016;48(7):1–11.

487. Mungall a J, Palmer S a, Sims SK, Edwards C a, Ashurst JL, Wilming L, et al. The

DNA sequence and analysis of human chromosome 6. Nature. 2003 Oct 23;425(6960):805–11.

488. Fehrmann RSN, Jansen RC, Veldink JH, Westra H-J, Arends D, Bonder MJ, et al. Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. PLoS Genet. 2011;7(8):e1002197.

489. Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A, et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat Genet. 2012;44(5):502–10.

490. Grimwood J, Gordon L a, Olsen A, Terry A, Schmutz J, Lamerdin J, et al. The DNA sequence and biology of human chromosome 19. Nature. 2004;428(6982):529–35.

491. Yang S, Liu Y, Jiang N, Chen J, Leach L, Luo Z, et al. Genome-wide eQTLs and heritability for gene expression traits in unrelated individuals. BMC Genomics. 2014;15(1):13.

492. Wright F a, Sullivan PF, Brooks AI, Zou F, Sun W, Xia K, et al. Heritability and genomics of gene expression in peripheral blood. Nat Genet. 2014;46(5):430–7.

493. Brostoff J, Mowbray J, Kapoor A, Hollowell S, Rudolf M, Saunders K. 80% of patients with intrinsic asthma are homozygous for HLA W6. Is intrinsic asthma a recessive disease? Lancet. 1976;2(7991):872–3.

494. Bruce CA, Bias WB, Norman PS, Lightenstein LM, Marsh DG. Studies of HLA antigen frequencies, IgE levels, and specific allergic sensitivities in patients having ragweed hayfever, with and without asthma. Clin Exp Immunol. 1976;25(1):67–72.

495. Kontakioti E, Domvri K, Papakosta D, Daniilidis M. HLA and asthma phenotypes/endotypes: A review. Hum Immunol. 2014;75(8):930–9.

496. Knowlden SA, Hillman SE, Chapman TJ, Patil R, Miller DD, Tigyi G, et al. Novel Inhibitory Effect of a Lysophosphatidic Acid 2 Agonist on Allergen-Driven Airway Inflammation. Am J Respir Cell Mol Biol. 2016;54(3):402–9.

497. Ackerman SJ, Park GY, Christman JW, Nyenhuis S, Berdyshev E, Natarajan V. Polyunsaturated lysophosphatidic acid as a potential asthma biomarker. J Allergy Clin Immunol. 2015;135(2 SUPPL. 1):AB151.

498. Park GY, Lee YG, Berdyshev E, Nyenhuis S, Du J, Fu P, et al. Autotaxin production of lysophosphatidic acid mediates allergic asthmatic inflammation. Am J Respir Crit Care Med. 2013;188(8):928–40.

499. Toews ML, Ustinova EE, Schultz HD. Lysophosphatidic acid enhances contractility of isolated airway smooth muscle. J Appl Physiol. 1997;83(4):1216–22.

500. Orrù S, Giuressi E, Carcassi C, Casula M, Contu L. Mapping of the major psoriasis-susceptibility locus (PSORS1) in a 70-Kb interval around the corneodesmosin gene (CDSN). Am J Hum Genet. 2005;76(1):164–71.

501. Teye K, Hamada T, Krol RP, Numata S, Ishii N, Matsuda M, et al. Homozygous deletion of six genes including corneodesmosin on chromosome 6p21.3 is associated with generalized peeling skin disease. J Dermatol Sci. 2014;75(1):36–42.

502. Kawakami H, Uchiyama M, Maeda T, Tsunoda T, Mitsuhashi Y, Tsuboi R. A Case of Inflammatory Generalized Type of Peeling Skin Syndrome Possibly Caused by a Homozygous Missense Mutation of CDSN. Case Rep Dermatol. 2014;6(3):232–8.

503. Israeli S, Zamir H, Sarig O, Bergman R, Sprecher E. Inflammatory peeling skin syndrome caused by a mutation in CDSN encoding corneodesmosin. J Invest Dermatol. 2011;131(3):779–81.

504. Palmer CN a, Irvine AD, Terron-Kwiatkowski A, Zhao Y, Liao H, Lee SP, et al. Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. Nat Genet. 2006;38(4):441–6.

505. Hudson TJ. Skin barrier function and allergic risk. Nat Genet. 2006;38(4):399–400.

506. Harly C, Guillaume Y, Nedellec S, Peigné C, Mönkkönen H, Li J, et al. Key implication of CD277 / Butyrophilin-3 (BTN3A) in cellular stress sensing by a major human γ δ T cell subset. Blood. 2012;120(11):2269–79.

507. Messal N, Mamessier E, Sylvain A, Celis-Gutierrez J, Thibult ML, Chetaille B, et al. Differential role for CD277 as a co-regulator of the immune signal in T and NK cells. Eur J Immunol. 2011;41(12):3443–54.

508. Le Page C, Marineau A, Bonza PK, Rahimi K, Cyr L, Labouba I, et al. BTN3A2 expression in epithelial ovarian cancer is associated with higher tumor infiltrating T cells and a better prognosis. PLoS One. 2012;7(6):e38541.

509. Al-Shibli KI, Donnem T, Al-Saad S, Persson M, Bremnes RM, Busund LT. Prognostic effect of epithelial and stromal lymphocyte infiltration in non-small cell lung cancer. Clin Cancer Res. 2008;14(16):5220–7.

510. Tada Y, Yamaguchi Y, Kinjo T, Song X, Akagi T, Takamura H, et al. The stem cell transcription factor ZFP57 induces IGF2 expression to promote anchorage-independent growth in cancer cells. Oncogene. 2014;34(6):752–60.

511. Kohda M, Hoshiya H, Katoh M, Tanaka I, Masuda R, Takemura T, et al. Frequent loss of imprinting of IGF2 and MEST in lung adenocarcinoma. Mol Carcinog. 2001;31(4):184–91.

512. Ozaki S, Kawahara E, Maenaka S, Hoang N, Oyama T, Imai M, et al. Distinct allelic expression patterns of imprinted IGF2 in adenocarcinoma and squamous cell

carcinoma of the lung. Oncol Lett. 2014;2561–4.

513. Takanami I, Imamuma T, Hashizume T, Kikuchi K, Yamamoto Y, Yamamoto T, et al. Insulin-like growth factor-II as a prognostic factor in pulmonary adenocarcinoma. J Surg Oncol. 1996;61(3):205–8.

514. Flomenberg N, Baxter-Lowe L a, Confer D, Fernandez-Vina M, Filipovich a, Horowitz M, et al. Impact of HLA class I and class II high resolution matching on outcomes of unrelated donor bone marrow transplantation: HLA-C mismatching is associated with a strong adverse effect on transplant outcome. Blood. 2004;104(0006–4971 LA–ENG PT–JOURNAL ARTICLE):1923–31.

515. Takemoto S, Port FK, Claas FHJ, Duquesnoy RJ. HLA matching for kidney transplantation. Hum Immunol. 2004;65(12):1489–505.

516. Chen X, Lu X, Chen J, Wu D, Qiu F, Xiong H, et al. Association of nsv823469 copy number loss with decreased risk of chronic obstructive pulmonary disease and pulmonary function in Chinese. Sci Rep. 2017;12(7):40060.

517. Wortham BW, Eppert BL, Motz GT, Flury JL, Orozco-Levi M, Hoebe K, et al. NKG2D mediates NK cell hyperresponsiveness and influenza-induced pathologies in a mouse model of chronic obstructive pulmonary disease. J Immunol. 2012;188(9):4468–75.

518. Telcian AG, Zdrenghea MT, Caramori G, Laza-Stanca V, Message SD, Kebadze T, et al. Soluble major histocompatibility complex class I-related chain B molecules are increased and correlate with clinical outcomes during rhinovirus infection in healthy subjects. Chest. 2014;146(1):32–40.

519. Silva D, Venihaki M, Guo WH, Lopez MF. Igf2 deficiency results in delayed lung development at the end of gestation. Endocrinology. 2006;147(12):5584–91.

520. Consensus M. American Thoracic Society American Thoracic Society / European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. Crit Care Med. 2002;165(2):277–304.

521. Tager AM, LaCamera P, Shea BS, Campanella GS, Selman M, Zhao Z, et al. The lysophosphatidic acid receptor LPA1 links pulmonary fibrosis to lung injury by mediating fibroblast recruitment and vascular leak. Nat Med. 2008;14(1):45–54.

522. Oikonomou N, Mouratis MA, Tzouvelekis A, Kaffe E, Valavanis C, Vilaras G, et al. Pulmonary autotaxin expression contributes to the pathogenesis of pulmonary fibrosis. Am J Respir Cell Mol Biol. 2012;47(5):566–74.

523. Montesi SB, Mathai SK, Brenner LN, Gorshkova I a, Berdyshev E V, Tager AM, et al. Docosatetraenoyl LPA is elevated in exhaled breath condensate in idiopathic pulmonary fibrosis. BMC Pulm Med. 2014;14(5):1–7.

524. Fairfax BP, Humburg P, Makino S, Naranbhai V, Wong D, Lau E, et al. Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression. Science (80-). 2014;343(6175):1246949.

525. Peters JE, Lyons PA, Lee JC, Richard AC, Fortune MD, Newcombe PJ, et al. Insight into Genotype-Phenotype Associations through eQTL Mapping in Multiple Cell Types in Health and Immune-Mediated Disease. PLoS Genet. 2016;12(3).

526. Delva E, Tucker DK, Kowalczyk AP. The Desmosome. 2009;1(2):17.

527. Al-Owain M, Wakil S, Shareef F, Al-Fatani A, Hamadah E, Haider M, et al. Novel homozygous mutation in DSP causing skin fragility-woolly hair syndrome: Report of a large family and review of the desmoplakin-related phenotypes. Clin Genet. 2011;80(1):50–8.

528. Awad MM, Calkins H, Judge DP. Mechanisms of disease: molecular genetics of arrhythmogenic right ventricular dysplasia/cardiomyopathy. Nat Clin Pr Cardiovasc Med. 2008;5(5):258–67.

529. Keith D, Armstrong B, McKenna KE, Purkis PE, Green KJ, Eady RAJ, et al. Haploinsufficiency of desmoplakin causes a striate subtype of palmoplantar keratoderma. Hum Mol Genet. 1999;8(1):143–8.

530. Whittock N V., Ashton GHS, Dopping-Hepenstal PJC, Gratian MJ, Keane FM, Eady RAJ, et al. Striate palmoplantar keratoderma resulting from desmoplakin haploinsufficiency. J Invest Dermatol. 1999;113(6):940–6.

531. Whittock N V., Wan H, Morley SM, Garzon MC, Kristal L, Hyde P, et al. Compound heterozygosity for non-sense and mis-sense mutations in desmoplakin underlies skin fragility/woolly hair syndrome. J Invest Dermatol. 2002;118(2):232–8.

532. Yang Z, Bowles NE, Scherer SE, Taylor MD, Kearney DL, Ge S, et al. Desmosomal dysfunction due to mutations in desmoplakin causes arrhythmogenic right ventricular dysplasia/cardiomyopathy. Circ Res. 2006;99(6):646–55.

533. Betancor MB, Almaida-Pagán PF, Sprague M, Hernández A, Tocher DR. Roles of selenoprotein antioxidant protection in zebrafish, Danio rerio, subjected to dietary oxidative stress. Fish Physiol Biochem. 2015;41(3):705–20.

534. Penglase S, Hamre K, Ellingsen S. Selenium prevents downregulation of antioxidant selenoprotein genes by methylmercury. Free Radic Biol Med. 2014;75:95–104.

535. Tinggi U. Selenium: Its role as antioxidant in human health. Environ Health Prev Med. 2008;13(2):102–8.

536. Kang M, Yoon CM, Kim BH, Lee C-M, Zhou Y, Sauler M, et al. Suppression of NLRX1 in chronic obstructive pulmonary disease. J Clin Invest. 2015;125(6):2458–

62.

537. Lei Y, Wen H, Ting JPY. The NLR protein, NLRX1, and its partner, TUFM, reduce type I interferon, and enhance autophagy. Autophagy. 2013;9(3):432–3.

538. Lei Y, Wen H, Yu Y, Taxman DJ, Zhang L, Widman DG, et al. The Mitochondrial Proteins NLRX1 and TUFM Form a Complex that Regulates Type I Interferon and Autophagy. Immunity. 2012;36(6):933–46.

539. Sato Y, Akitsu M, Amano Y, Yamashita K, Ide M, Shimada K, et al. The novel PAR-1-binding protein MTCL1 has crucial roles in organizing microtubules in polarizing epithelial cells. J Cell Sci. 2013;126(20):4671–83.

540. Sato Y, Hayashi K, Amano Y, Takahashi M, Yonemura S, Hayashi I, et al. MTCL1 crosslinks and stabilizes non-centrosomal microtubules on the Golgi membrane. Nat Commun. 2014;5:5266.

541. Wert SE, Yoshida M, LeVine a M, Ikegami M, Jones T, Ross GF, et al. Increased metalloproteinase activity, oxidant production, and emphysema in surfactant protein D gene-inactivated mice. Proc Natl Acad Sci U S A. 2000;97(11):5972–7.

542. Blake JA, Eppig JT, Kadin JA, Richardson JE, Smith CL, Bult CJ, et al. Mouse Genome Database (MGD)-2017: community knowledge resource for the laboratory mouse. Nucleic Acids Res. 2017;45(Databse issue):D723–D729.

543. Ou CY, Chen CZ, Hsiue TR, Lin SH, Wang JY. Genetic variants of pulmonary SP-D predict disease outcome of COPD in a Chinese population. Respirology. 2015;20(2):296–303.