



Marqueurs génétiques influençant l'expression des gènes dans les poumons et susceptibilité à la maladie pulmonaire obstructive chronique

Mémoire

Maxime Lamontagne

Maîtrise en Médecine Expérimentale
Maître ès sciences (M. Sc.)

Québec, Canada

© Maxime Lamontagne, 2013

Résumé

La maladie pulmonaire obstructive chronique (MPOC) se caractérise par une obstruction non complètement réversible des voies aériennes. De récentes études de criblage génomique par association ont identifié quatre régions chromosomiques associées à la MPOC. Par contre, les mécanismes moléculaires responsables de ces associations génétiques demeurent inconnus. Dans le cadre de cette étude, nous avons obtenu les profils d'expression des gènes de tissus pulmonaires non-tumoraux et les génotypes de 1,2 million de variations génétiques pour ces mêmes patients. Les analyses furent effectuées sur 1111 sujets provenant de trois cohortes. Des marqueurs génétiques influençant l'expression des gènes dans les poumons, c.-à-d. « expression Quantitative Trait Loci » (eQTLs) pulmonaires, furent identifiés dans les régions chromosomiques d'intérêts. Les résultats de ce mémoire démontrent que *HHIP* serait le gène causal dans la région 4q31, alors que les évidences pointent vers les gènes *FAM13A* et *EGLN2* pour les régions 4q22 et 19q13, respectivement.

Abstract

Chronic obstructive pulmonary disease (COPD) is characterized by airflow obstruction that is not fully reversible. Recent genome-wide association studies have identified four susceptibility loci robustly associated with COPD. However, the genetic mechanisms mediating the risk within these loci remain to be found. In this study, genome-wide gene expression profiles of non-tumor lung specimens and blood-DNA from the same patients were genotyped for 1,2 million SNPs. The analyses were performed on 1111 subjects from three cohorts. Genetics variations influencing gene expression levels in lung samples, i.e. lung expression quantitative trait loci (eQTLs), were identified in the COPD susceptibility regions (4q22, 4q31, 19q13). The results of this thesis demonstrated that *HHIP* is the most likely causal gene at 4q31, while the evidences supported the contribution of the *FAM13A* and *EGLN2* genes at 4q22 and 19q13, respectively.

Table des matières

Résumé.....	iii
Abstract.....	v
Table des matières.....	vii
Liste des tableaux.....	ix
Liste des figures.....	xi
Liste des abréviations et des sigles.....	xiii
Remerciements.....	xv
Avant-propos.....	xvii
Chapitre 1 : Introduction.....	1
1. La MPOC.....	2
1.1. Définition.....	2
1.2. Épidémiologie et prévalence.....	2
1.2. Conséquence de la MPOC: Présentation clinique et évaluation.....	3
1.2.1. Symptômes.....	3
1.2.2. Diagnostic et classification de la sévérité de la MPOC.....	3
1.3. Maladies coexistantes dans la MPOC.....	4
1.3.1. L'emphysème.....	5
1.3.1.1. Symptômes cliniques.....	6
1.3.1.2. Types d'emphysèmes.....	6
1.3.1.3. La pathophysiologie.....	6
1.3.2. La bronchite chronique.....	7
1.3.2.1. Symptômes cliniques.....	8
1.3.2.2. La pathophysiologie.....	8
1.4. Les causes de la MPOC.....	8
1.4.1. Tabagisme.....	8
1.4.2. Déficience en alpha-1 antitrypsine.....	10
1.4.3. Autres causes.....	11
1.5. Traitements.....	11
1.5.1. Traitements non pharmacologiques.....	11
1.5.2. Traitements pharmacologiques.....	12
1.5.3. Traitement de la déficience en alpha-1 antitrypsine.....	12
2. Génétique de la MPOC.....	14
2.1. Terminologie génétique.....	14
2.1.1. Génotype.....	14
2.1.2. expression Quantitative Trait Loci (eQTLs).....	15
2.2. La génétique de la MPOC.....	16
2.2.1. Déficience en alpha-1 antitrypsine.....	16
2.2.1.1. Diagnostique de la déficience.....	18
2.2.2. Les études par gènes candidats.....	19
2.2.3. Les criblages génomiques par association (GWAS).....	22
2.3. Les limitations des études génétiques précédentes.....	25
3. Matériels et méthodes.....	28
3.1. Sujets de recherche.....	28
3.2. Approches génomiques utilisées.....	28

3.2.1. Expression des gènes.....	28
3.2.2. Génotypage.....	29
3.3. Outils bioinformatiques.....	29
3.3.1. Linux.....	29
3.3.2. PLINK.....	29
3.3.3. Affymetrix Power Tools.....	30
3.3.4. STRUCTURE.....	30
3.3.5. R.....	31
3.4. Analyses Statistiques.....	31
4. Objectif et Hypothèses.....	32
4.1. Objectif.....	32
4.2. Hypothèses.....	32
Chapitre 2 : Refining susceptibility loci of chronic obstructive pulmonary disease with lung eQTLs...	33
Chapitre 3 : Conclusion.....	71
Bibliographie.....	79

Liste des tableaux

Chapitre 1

Tableau 1: Classification de la MPOC selon le test de spirométrie.....	4
Tableau 2: Mutations dans le gène <i>SERPINA1</i>	17
Tableau 3: Gènes et régions chromosomiques associées à la MPOC et aux phénotypes reliés par GWAS.....	23

Chapitre 2

Table 1: Clinical characteristic of patients that passed gene expression and genotyping quality control filters.....	49
Table 2: SNPs associated with COPD in previous GWAS.....	50
Table S1: Significant eQTLs at the 4q22 locus in the Laval dataset and replication in UBC and Groningen.....	60
Table S2: Significant eQTLs at the 4q31 locus in the Laval dataset and replication in UBC and Groningen.....	62
Table S3: Significant eQTLs at the 19q13 locus in the Laval dataset and replication in UBC and Groningen.....	64

Liste des figures

Chapitre 1

Figure 1: Représentation des différences entre un poumon en santé et celui d'une personne souffrant d'emphysème.....	5
Figure 2: Représentation des différences entre les bronches d'une personne en bonne santé et ceux d'une personne souffrant de bronchite chronique	7
Figure 3: Polymorphisme Nucléotidique Simple (SNP).....	14
Figure 4: Exemple d'un eQTL.....	15
Figure 5: Représentation schématisée des cis-eQTLs et des trans-eQTLs sur un même chromosome.....	16
Figure 6: Gènes candidats de la MPOC.....	20
Figure 7: Gènes candidats associés à la MPOC et aux phénotypes reliés à cette maladie.....	22

Chapitre 2

Figure 1: Lung eQTLs on 4q22 in the Laval dataset.....	51
Figure 2: Boxplots of lung gene expression levels for <i>PPM1K</i> according to genotype groups for SNP rs17013978.....	52
Figure 3: Linkage disequilibrium plot of selected SNPs on the 4q22 locus in the 1000 Genome Project.....	53
Figure 4: Boxplots of lung gene expression levels for <i>FAM13A</i> according to genotype groups for SNP rs2045517.....	54
Figure 5: Lung eQTLs on 4q31 in the Laval dataset.....	55
Figure 6: Boxplots of lung gene expression levels for <i>BC029578</i> according to genotype groups for SNP rs7667092.....	56
Figure 7: Boxplots of lung gene expression levels for <i>HHIP</i> according to genotype groups for SNP rs1828591.....	57
Figure 8: Lung eQTLs on 19q13 in the Laval dataset.....	58
Figure 9: Boxplots of lung gene expression levels for <i>EGLN2</i> according to genotype groups for SNP rs4803369.....	59
Figure S1: Linkage disequilibrium plot of significant SNPs on the 4q22 locus in the 1000 Genome Project.....	68
Figure S2: Linkage disequilibrium plot of significant SNPs on the 4q31 locus in the 1000 Genome Project.....	69
Figure S3: Linkage disequilibrium plot of significant SNPs on the 19q13 locus in the 1000 Genome Project.....	70

Liste des abréviations et des sigles

ACLA	Anticholinergiques à longue durée d'action
ADN	Acide désoxyribonucléique
APT	Affymetrix power tools
ARNm	Acide ribonucléique messenger
BALA	β 2-adrénergiques à longue durée d'action
BDCA	Bronchodilatateurs à courte durée d'action
BPCO	Bronchopneumopathie chronique obstructive
CRF	Capacité résiduelle fonctionnelle
CRIUCPQ	Centre de recherche de l'institut universitaire de cardiologie et de pneumologie de Québec
CSI	Corticostéroïdes en inhalation
CVF	Capacité vitale forcée
eQTLs	expression quantitative trait loci
FEV1	Forced expiratory volume in 1 seconde
FRQS	Fonds de recherche du Québec – Santé
FVC	Forced vital capacity
GEO	Gene expression omnibus
GOLD	Global initiative for chronic obstructive lung disease

GRNG	Université de Groningen
GWAS	Criblages génomiques par association
IMC	Indice de masse corporelle
IUCPQ	Institut universitaire de cardiologie et de pneumologie de Québec
LD	Déséquilibre de liaison
MPOC	Maladie pulmonaire obstructive chronique
pre-BD	pré-bronchodilatateur
SNP	Polymorphisme nucléotidique simple
UBC	Université de la Colombie-Britannique
VEMS	Volume expiratoire maximal seconde

Remerciements

Merci au Dr Yohan Bossé de m'avoir accueillie dans son laboratoire et qui a vu en moi un potentiel que je ne connaissais pas. Merci à Nathalie Gaudreault pour ses explications à mes questions nombreuses (sans oublier la nourriture et les remboursements). Merci également à tous les autres membres du laboratoire que j'ai côtoyé quotidiennement pendant ces deux années : Justin Dang Uy Nguyen, Jean-Christophe Bérubé, Laura Sbara, Cyndi Henry, Emilie Lavoie-Charland, Valérie Ducharme, Sandra Guauque-Olarte et les nombreux stagiaires d'été. Je tiens à remercier tous les collaborateurs avec qui j'ai travaillé au cours de ces années. Finalement, merci à ma famille et mes amis qui m'ont permis de conserver un semblant de vie sociale lors de cette période charnière de ma vie.

J'aimerais bien écrire un texte beaucoup plus long pour vous exprimer ma gratitude, mais mes talents dans ce domaine sont limités. Je suis certain qu'au fond de vous, vous êtes conscient que je vous serai éternellement reconnaissant.

Avant-propos

Le présent mémoire est constitué d'une introduction sur la maladie pulmonaire obstructive chronique (MPOC) et la génétique, d'un manuscrit scientifique et d'une conclusion générale sur les marqueurs génétiques précédemment associés à la MPOC influençant l'expression des gènes dans les poumons. Le manuscrit sera soumis à la revue PLoS One lorsque tous les collaborateurs auront donné leur aval.

Les résultats rapportés dans ce mémoire proviennent des données recueillies par le « Merck-Laval-UBC-Groningen Lung eQTL consortium ». Dr Yohan Bossé est professeur à l'Université Laval et membre du centre de recherche de l'Institut universitaire de cardiologie et de pneumologie de Québec (CRIUCPQ). Il a participé à la planification du projet et à l'écriture de l'article. Dr Michel Laviolette et Dr Christian Couture sont tous deux médecins à l'Institut universitaire de cardiologie et de pneumologie de Québec (IUCPQ) et chercheurs au CRIUCPQ. Ils ont participé à la création de la cohorte de Laval. Dr Peter D. Paré et Dr Don D. Sin sont chercheurs au UBC James Hogg Research Center, Providence Heart and Lung Institute, St. Paul's Hospital de Vancouver. Ils sont également professeurs à l'Université de la Colombie-Britannique. Ils ont participé à la création de la cohorte de UBC. Dr Dirkje S. Postma et Dr Wim Timens sont professeurs à l'Université de Groningen au Pays-Bas. Ils ont participé à la création de la cohorte de Groningen. Dr David Nickle est chercheur pour la compagnie Merck & Co. Inc.. Il a participé au génoypage et à la quantification de l'expression des gènes par biopuces d'ADN. Tous les co-auteurs ont participé à la révision du manuscrit. À titre de premier auteur, j'ai effectué les analyses statistiques et bioinformatiques et rédigé le manuscrit sous la supervision du Dr Bossé. Aucune modification du manuscrit n'a été effectuée dans le mémoire.

Chapitre 1 : Introduction

1. La MPOC

1.1. Définition

La maladie pulmonaire obstructive chronique (MPOC) est une maladie qui affecte les voies respiratoires et les poumons. Elle se caractérise par une réduction lente et progressive des fonctions respiratoires¹. Contrairement à l'asthme, la limitation de la capacité respiratoire associée à la MPOC n'est pas complètement réversible. La réduction de la capacité respiratoire a comme origine une obstruction des voies respiratoires, des bronches, et/ou une destruction permanente du tissu pulmonaire, les alvéoles. La MPOC se compose essentiellement de deux maladies pulmonaires différentes; la bronchite chronique et l'emphysème². Une personne atteinte de MPOC peut souffrir de l'une ou l'autre de ces maladies, mais une combinaison des deux maladies est le plus souvent présente². La MPOC porte également le nom de bronchopneumopathie chronique obstructive (BPCO).

1.2. Épidémiologie et prévalence

La MPOC est une maladie dont la prévalence augmente avec l'âge. Elle se développe habituellement chez les fumeurs de longue date³. Les symptômes de la maladie apparaissent rarement avant l'âge de 55 ans. Une méta-analyse publiée en 2006 indique que la prévalence mondiale de la MPOC est de 8,9 %⁴. La prévalence est plus élevée chez les hommes (11,0 %) que chez les femmes (5,0 %). Selon cette même étude, l'Amérique est le continent le moins touché avec une prévalence de 4,5 % tandis que le sud-est de l'Asie est la région la plus touchée avec une prévalence de 12,5 %. D'ici 2030, il est estimé que la MPOC deviendra la troisième cause de décès dans le monde⁵. Elle sera alors la cause de 7,8 % des décès à l'échelle mondiale. En 2009, 4,2 % des Canadiens âgés de 35 ans et plus souffraient de MPOC⁶. Selon cette source, les femmes (4,5 %) étaient légèrement plus atteintes que les hommes (4,0 %). Après l'âge de 75 ans, une plus grande proportion de la population est atteinte. Dans cette tranche d'âge, les hommes (10,8 %) sont plus touchés que les femmes (8,4 %)⁶. En 2004, la MPOC fût la

cause de 5,152 décès chez les hommes et de 4,455 chez les femmes au Canada⁷. Les décès causés par la MPOC (9,607) représentent 4 % de tous les décès de l'année 2004 au Canada⁷. En 2008, le fardeau économique de la MPOC a été estimé à 1,5 milliard de dollars au Canada⁸.

1.2. Conséquence de la MPOC: Présentation clinique et évaluation

1.2.1. Symptômes

Les symptômes de la MPOC sont vastes et très handicapant pour les personnes qui en souffrent. Le symptôme le plus important chez les personnes MPOC est la dyspnée⁹. Au début de la maladie, cette difficulté respiratoire est présente seulement à l'effort, mais peut apparaître au repos au cours de l'évolution de la maladie. Les autres symptômes pouvant être présents chez les personnes malades sont une abondance de mucus dans les voies respiratoires, une toux persistante, une fatigue générale, des infections pulmonaires fréquentes, une douleur thoracique, une perte de masse musculaire et une perte de poids. Ces symptômes ne sont pas présents chez tous les patients MPOC. Leur présence dépend entre autre de la gravité de la maladie et de la cause de celle-ci (emphysème/bronchite chronique). Chez certains patients, la toux chronique et la production de mucus précèdent la réduction des fonctions respiratoires. Par contre, pour d'autres patients, les limitations respiratoires peuvent survenir sans toux ou production de mucus.

1.2.2. Diagnostic et classification de la sévérité de la MPOC

Le diagnostic de la MPOC s'effectue à l'aide d'un test de spirométrie. Deux variables mesurées lors de ce test sont utilisées afin d'établir le diagnostic de la MPOC: 1) le Volume Expiratoire Maximal Seconde (VEMS) et 2) 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. Les valeurs de VEMS et CVF sont habituellement obtenues suivant l'inhalation d'un bronchodilatateur. Le diagnostic sera par la suite dicté par

les standards GOLD (*Global Initiative for Chronic Obstructive Lung Disease*)¹. Ces critères permettent de déterminer dans un premier temps si une personne souffre de MPOC, puis si c'est le cas, dans un deuxième temps classer le stade de la maladie. Pour poser un diagnostic de MPOC, le rapport de VEMS/CVF doit être inférieur à 0,7 de la valeur prédite post-bronchodilatateur. Le VEMS est par la suite utilisé pour déterminer le stade de la maladie¹. Le Tableau 1 présente les critères de spirométrie couramment utilisés afin de classer la sévérité de la MPOC en quatre stades. À noter que l'évaluation de la sévérité de la MPOC n'est pas uniquement basée sur ces simples critères spirométriques. Le stade final de la maladie provient d'un jugement clinique basé également sur les symptômes et les complications du patient.

Tableau 1: Classification de la MPOC selon le test de spirométrie.

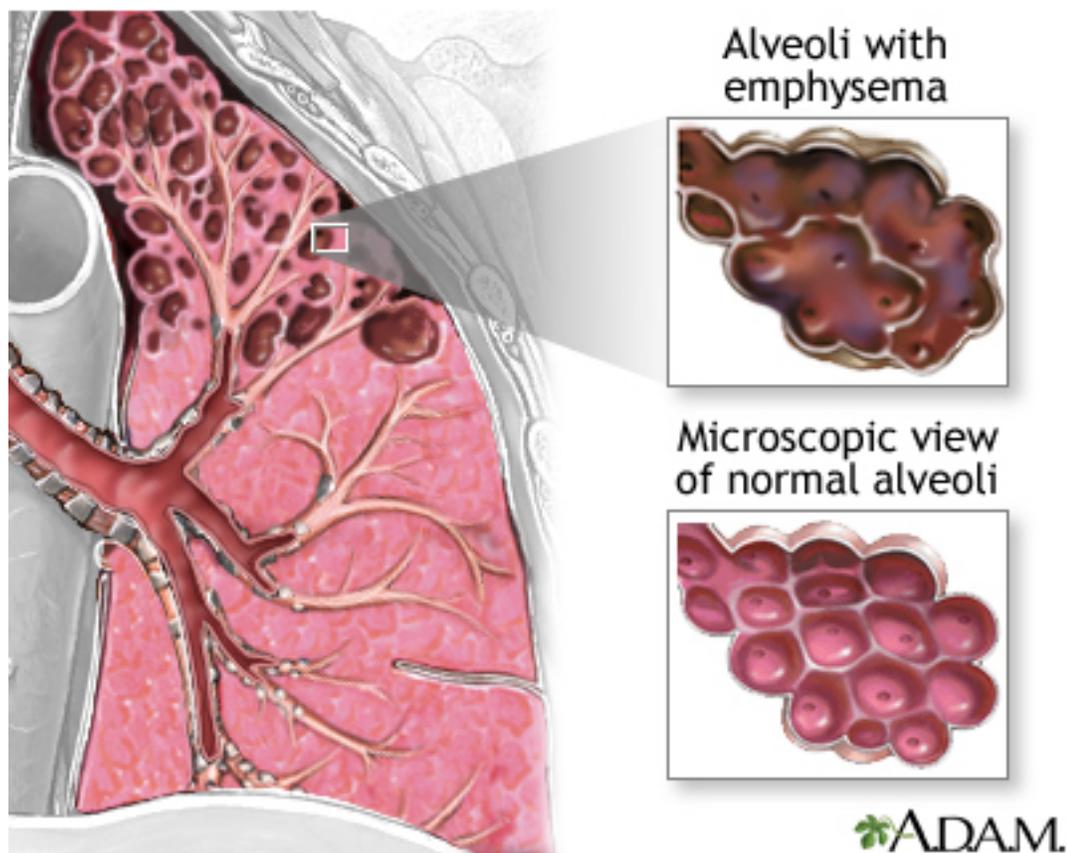
Stade	VEMS/CVF (post bronchodilatateur)	VEMS (post bronchodilatateur)
1 : Léger	VEMS/CVF < 0,70	VEMS ≥ 80% de la valeur prédite
2 : Modéré	VEMS/CVF < 0,70	50% ≤ VEMS < 80% de la valeur prédite
3 : Sévère	VEMS/CVF < 0,70	30% ≤ VEMS < 50% de la valeur prédite
4 : Très Sévère	VEMS/CVF < 0,70	VEMS < 30% de la valeur prédite

1.3. Maladies coexistantes dans la MPOC

Tel que mentionné à la section 1.1.1, la MPOC se compose essentiellement de deux maladies pulmonaires; la bronchite chronique et l'emphysème². Bien que ces deux maladies n'affectent pas les mêmes tissus, les bronches pour la bronchite chronique et les alvéoles pour l'emphysème, elles conduisent toutes les deux à une réduction de la capacité respiratoire.

1.3.1. L'emphysème

L'emphysème est une maladie pulmonaire se caractérisant par une destruction du tissu pulmonaire (Figure 1)¹⁰. Cette destruction progressive évolue dans le temps jusqu'à créer des espaces où les échanges gazeux sont impossibles. La principale cause conduisant au phénotype d'emphysème est le tabagisme. De plus, une déficience en alpha-1 antitrypsine aura comme conséquence de provoquer l'apparition de l'emphysème même chez des patients non fumeurs ou chez des fumeurs plus jeune vers l'âge de 40 ans^{11,12}.



Copyright © 2012, A.D.A.M., Inc.

Figure 1 : Représentation des différences entre un poumon en santé (coin inférieur droit) et celui d'une personne souffrant d'emphysème (coin supérieur droit). Les sacs alvéolaires de la personne malade sont partiellement détruits affectant ainsi les échanges gazeux.

1.3.1.1. Symptômes cliniques

Le principal symptôme clinique de l'emphysème est une dyspnée présente surtout à l'effort. Le poumon présentera de l'hyperinflation et le patient peut ressentir un inconfort thoracique¹³. Par contre, il n'y aura pas de surproduction de mucus ni de toux chronique. La capacité résiduelle fonctionnelle (CRF) du patient, i.e. le volume d'air contenu le poumon après une expiration normale, sera augmentée puisque celui-ci aura de la difficulté à vider ses poumons. Ceci est la cause de la rétention gazeuse qui se traduit par un volume résiduel augmenté.

1.3.1.2. Types d'emphysèmes

Il existe deux types d'emphysèmes soit l'emphysème centribulaire et panlobulaire¹⁴. L'emphysème centribulaire est principalement développé chez les fumeurs. Suivant l'exposition à la fumée, les particules nocives sont piégées dans les alvéoles pulmonaires. Une inflammation localisée dans l'alvéole se produit et causant ainsi des dommages à la structure des alvéoles. Graduellement, les alvéoles se rompent et fusionnent pour créer de grandes poches d'air qui ne participent plus aux échanges gazeux. Cette destruction alvéolaire se produit principalement dans les lobes supérieurs et se concentre dans les portions centrales des lobules secondaires¹⁵. D'autre part, l'emphysème panlobulaire est une caractéristique d'une déficience en alpha-1 antitrypsine¹⁶. Le poumon est alors détruit de façon plus égale dans toutes ses structures. Par contre, les lobes inférieurs sont principalement touchés. Cette déficience est présente chez seulement 1 à 5 % des patients MPOC¹⁷.

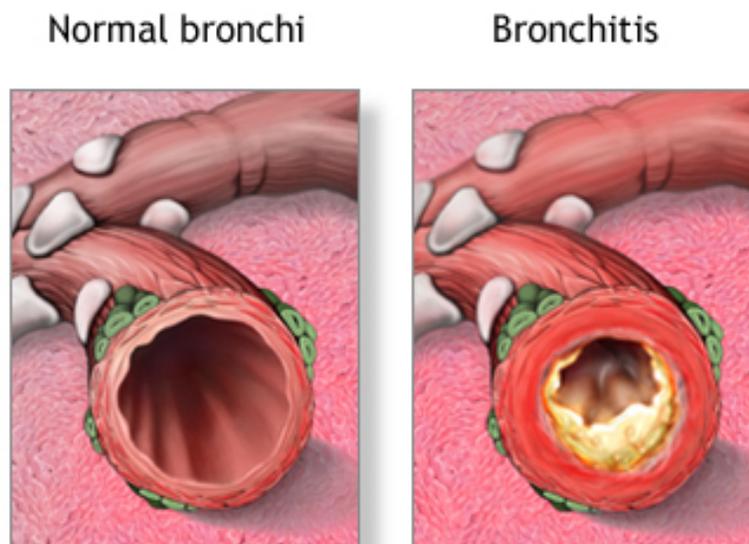
1.3.1.3. La pathophysiologie

Les hypothèses physiopathologiques sont nombreuses. L'hypothèse la plus fréquemment citée est le déséquilibre protéase/antiprotéase qui serait à l'origine de la destruction et de l'élargissement des alvéoles¹⁸. Ce déséquilibre a pour origine la libération de protéases par les cellules inflammatoires telle que les neutrophiles et les macrophages qui sont nombreux dans l'inflammation associée à la MPOC. Ces cellules présentes en grande quantité vont être responsable de la

destruction des alvéoles¹⁰. Les cellules inflammatoires vont également être responsables de l'inflammation du poumon. Les échanges gazeux sont altérés suite à une réduction de la surface de contact alvéolaire. Les autres hypothèses physiopathologiques sont entre autres le stress oxydant et l'apoptose.

1.3.2. La bronchite chronique

La bronchite chronique est une maladie pulmonaire se caractérisant par une obstruction des bronches respiratoires². Cette obstruction peut être occasionnée par une surproduction de mucus, une inflammation bronchite et/ou une hyperréactivité bronchique (Figure 2). Elle a comme conséquence de rendre la respiration plus difficile (dyspnée). La cause la plus commune de la bronchite chronique est le tabagisme.



 ADAM.

Copyright © 2012, A.D.A.M., Inc.

Figure 2 : Représentation des différences entre les bronches d'une personne en bonne santé (gauche) et ceux d'une personne souffrant de bronchite chronique (droite). On peut voir que l'espace existant pour permettre le déplacement de l'air dans les poumons atteints est grandement réduit.

1.3.2.1. Symptômes cliniques

La bronchite chronique est caractérisée par une toux persistante et une surproduction de mucus. Ces deux premiers symptômes cliniques doivent être présents au moins trois mois par année pour deux années consécutives pour poser un diagnostic de bronchite chronique. Les personnes atteintes de la maladie souffriront également de dyspnée et parfois de fatigue générale et/ou de douleur thoracique.

1.3.2.2. La pathophysiologie

La bronchite chronique a plusieurs effets sur les bronches. Il y a présence d'une augmentation du nombre de cellules épithéliales glandulaires¹⁹ et également un accroissement des glandes sous-muqueuses dans les grandes voies aériennes²⁰. Ces modifications auront pour effet d'augmenter les sécrétions bronchiques. De plus, les muscles lisses deviennent plus rigides et inflammés²¹. Les parois bronchiques deviennent donc plus épaisses limitant ainsi le flux respiratoire. Tous ces symptômes combinés auront pour effet de limiter la circulation de l'air dans les voies aériennes.

1.4. Les causes de la MPOC

Tel que mentionné plus haut, le tabagisme est la principale cause pouvant mener à la MPOC. Par contre, seulement 15 à 20 % des fumeurs développent cette maladie¹. Certaines prédispositions génétiques semblent donc jouer un rôle dans la maladie²²⁻²⁴.

1.4.1. Tabagisme

Le tabagisme est de loin le principal facteur conduisant au développement de la MPOC. Il est responsable d'environ 85 à 90 % des cas. La fumée de cigarette contient plus de 5000 agents chimiques dont plusieurs ayant un effet néfaste pour la santé²⁵. Les effets néfastes de cette habitude de vie peuvent laisser des traces moléculaires dans les poumons même plusieurs années après l'arrêt tabagique²⁶. Le tabac cause une destruction des alvéoles pulmonaires. Pour ce

faire, la fumée de cigarette enclenche diverses réactions. L'activation de certaines cellules de l'immunité innée²⁷, le stress oxydant^{28,29} et une réponse immunitaire adaptative³⁰ sont trois réactions responsables de la détérioration des tissus pulmonaires. De plus, la fumée de cigarette a pour effet de détruire les cils vibratiles présents dans les voies aériennes supérieures. Cette destruction de cils a pour effet de rendre le système respiratoire plus propice aux infections virales.

L'immunité innée se caractérise par un nombre important de cellules qui constituent la première ligne de défense du système immunitaire. Dans le système pulmonaire, on compte un grand nombre de types cellulaires. Les éosinophiles, les neutrophiles, les macrophages et les cellules tueuses naturelles en sont quelques exemples. Ces cellules ont comme fonction d'éliminer les agents infectieux, de se débarrasser des cellules mortes et de réparer les dommages causés au tissu pulmonaire. Il existe un équilibre important entre la production cellulaire de protéases et d'antiprotéases. Un déséquilibre de ce système semble être présent chez les fumeurs souffrant de MPOC¹⁸. Un exemple de l'importance de cet équilibre est le résultat d'une déficience en alpha-1 antitrypsine. Cette enzyme est une anti-protéase importante dans le poumon et les personnes souffrant d'une déficience peuvent souffrir de MPOC sans aucune consommation tabagique^{11,12}. La présentation clinique d'un déficit en alpha-1 antitrypsine est l'apparition à un plus jeune âge, soit vers l'âge de 40-45 ans, d'une MPOC sévère chez les fumeurs.

Le stress oxydant est une réaction provoquée par la fumée de cigarette. Cette réaction a pour effet d'activer le recrutement de cellules inflammatoires et ainsi causer une inflammation des voies aériennes et des poumons²⁹. De plus, le stress oxydant peut parfois causer des dommages à l'ADN et ainsi mener diverses cellules à l'apoptose²⁸. Cela contribuera à la destruction des cellules pulmonaires tout en limitant leur régénération.

La réponse immunitaire adaptative est responsable des symptômes persistants après l'arrêt tabagique. Normalement, les cellules du système immunitaire se retirent lors de la fin de l'agression. Par contre, cette réversibilité n'est pas observée instantanément chez les sujets atteints de MPOC³¹. La réaction

immunitaire innée semble activer la réponse immunitaire adaptative même après la fin de l'agression³⁰. La réponse adaptative ne se comporte pas adéquatement en attaquant les cellules saines chez les sujets MPOC conduisant ultimement à une destruction pulmonaire.

1.4.2. Déficience en alpha-1 antitrypsine

L'alpha-1 antitrypsine est une anti-protéase sécrétée par le foie et qui a comme fonction d'inhiber l'élastase produite par les cellules inflammatoires^{17,32}. Une déficience en alpha-1 antitrypsine peut conduire à l'emphysème. Cette déficience est le fruit de mutations génétiques dans le gène *SERPINA1* localisé sur le chromosome 14q32. Tel que mentionné plus haut, cette déficience est présente chez 1 à 5 % des patients MPOC¹⁷.

Plusieurs variations génétiques ont été identifiées dans le gène *SERPINA1*³³. Le génotype le plus commun causant la déficience est le variant Z. Cette variation génétique consiste à un changement d'un acide glutamique pour une lysine à la position 242 de la protéine (Glu242Lys). De plus, une alanine doit être présente à la position 213 (Ala213). La deuxième variation génétique la plus courante pouvant mener à ce phénotype est le variant S. Ce variant est causé par un changement d'un acide glutamique pour une valine en position 264 (Glu264Val) et de la présence d'une valine en position 213 (Val213). Cependant, le variant S a beaucoup moins d'impact sur l'expression de l'alpha-1 antitrypsine contrairement au variant Z. En plus des variants Z et S, un très grand nombre de variants rares ont été identifiés à ce jour³³. Le variant Z est le mieux caractérisé. Le cas le plus ancien répertorié jusqu'à présent a été retrouvé chez une fillette de six ans qui fut congelée pendant 800 ans en Alaska³⁴. Des études récentes ont estimé que cette mutation aurait apparu il y a 2360 ans en Suède et il y a 2900 ans en Lettonie³⁵.

Le diagnostic d'une déficience en alpha-1 antitrypsine s'effectue initialement en dosant la présence de la protéine dans le sang³⁶. Les personnes habitant en Europe du nord sont plus susceptibles de souffrir de cette déficience³⁵. La fréquence du génotype le plus à risque, le génotype ZZ, est présent dans 0,05 % de la population dans plusieurs de ces pays³⁷.

1.4.3. Autres causes

La pollution atmosphérique peut également contribuer au développement de la MPOC. La pollution peut être présente en milieu de travail (poussière, produits chimiques) ou dans les grandes villes polluées (smog)⁹. Des habitations mal ventilées et l'utilisation de la biomasse pour la cuisson sont aussi des facteurs de risque pour la MPOC, particulièrement dans les pays en développement³⁸. Outre la déficience en alpha-1 antitrypsine, d'autres facteurs génétiques sont suspectés d'avoir un rôle dans le développement de la maladie et seront traités à la section 2.

1.5. Traitements

À ce jour, il n'existe aucune solution pour traiter la MPOC. Les traitements existants servent à ralentir la progression de la maladie, améliorer la qualité de vie des personnes souffrantes, réduire les effets néfastes de la maladie et diminuer le taux de mortalité. L'arrêt de toute consommation tabagique est un élément central de l'intervention médicale³⁹.

1.5.1. Traitements non pharmacologiques

Il existe plusieurs stratégies ne nécessitant aucune prise de médicament qui peuvent aider à mieux vivre avec la MPOC voir même améliorer la survie. Tout d'abord, la cessation tabagique est hautement recommandée chez les fumeurs. Tel que mentionné dans la section 1.4.1, le tabagisme est la principale cause menant au développement de la maladie. L'arrêt tabagique aura comme principal effet de ralentir la progression de la maladie en réduisant l'apport de produits néfastes dans les poumons³⁹. Les personnes atteintes de MPOC sont plus susceptibles aux infections respiratoires. La vaccination antigrippale est donc également recommandée surtout chez les personnes âgées⁴⁰. Il existe également certains programmes d'exercices pour aider les patients à développer et maintenir leur masse musculaire pour ainsi faciliter les activités quotidiennes^{41,42}. La réadaptation pulmonaire est également un excellent moyen de réduire l'essoufflement⁴³. L'oxygénothérapie à domicile peut aussi s'avérer nécessaire dans certains cas sévères de la maladie⁴⁴. Il y a donc trois interventions non pharmacologiques qui

permettent de réduire la mortalité liée à la MPOC : l'arrêt tabagique, l'oxygénothérapie chronique dans le cas sévère avec hypoxémie et la réadaptation.

1.5.2. Traitements pharmacologiques

Les traitements disponibles sur le marché ont majoritairement pour but de soulager la dyspnée. Actuellement, aucun médicament ne permet de renverser la diminution accélérée des fonctions respiratoires⁴⁵⁻⁴⁸. Il existe une gamme de médicaments qui sont prescrits selon la sévérité de la maladie et les antécédents médicamenteux. Pour les cas de MPOC léger, la première classe de médicament prescrit sont les bronchodilatateurs à courte durée d'action (BDCA). Cette classe de médicament est recommandée lors de difficultés respiratoires passagères pour détendre les bronches. Lorsque les BDCA s'avèrent insuffisants, une combinaison d'anticholinergiques à longue durée d'action (ACLA) et de BDCA ou une combinaison de β 2-adrénergiques à longue durée d'action (BALA) et de BDCA peuvent être utilisées. Ces médicaments ont comme fonction de prévenir les bronchospasmes à court et à long terme. Pour les cas plus sévère 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 β 2-adrénergique à longue durée d'action (CSI/BALA) et la théophylline peuvent aussi être employés.

1.5.3. Traitement de la déficience en alpha-1 antitrypsine

À l'heure actuelle, il n'existe aucun traitement pour rétablir des niveaux normaux de protéines. Par contre, des études ont tenté de démontrer l'efficacité d'un traitement substitutif d'alpha-1 antitrypsine afin de ralentir la progression de la maladie. Les résultats de ces études semblent montrer une amélioration pour certains paramètres: mortalité, déclin des fonctions respiratoires, densité pulmonaire¹⁷. Par contre, d'autres paramètres n'ont subi aucune amélioration lors

des traitements dont les exacerbations et la qualité de vie. De plus, plusieurs aspects cliniques n'ont pas encore fait l'objet d'études avec ce traitement tel la dyspnée, la capacité physique et les hospitalisations. Les experts canadiens en la matière s'accordent pour dire que le traitement devrait être offert aux patients non-fumeurs et ex-fumeurs atteint de MPOC souffrants de la déficience qui reçoivent des traitements pharmacologiques et non-pharmacologiques¹⁷. Cependant, des études supplémentaires devront être réalisées pour mieux comprendre l'effet de cette thérapie.

2. Génétique de la MPOC

2.1. Terminologie génétique

2.1.2. Génotype

Les polymorphismes nucléotidique simple (SNPs) sont des variations génétiques d'une seule paire de base observées dans la population (Figure 3). Ces variations génétiques sont très nombreuses dans le génome humain⁴⁹. Elles font partie du patrimoine génétique de chaque personne et certaines de ces variations sont responsables des différences morphologiques observées entre les individus, mais également des différences au niveau de la susceptibilité à de nombreuses maladies.

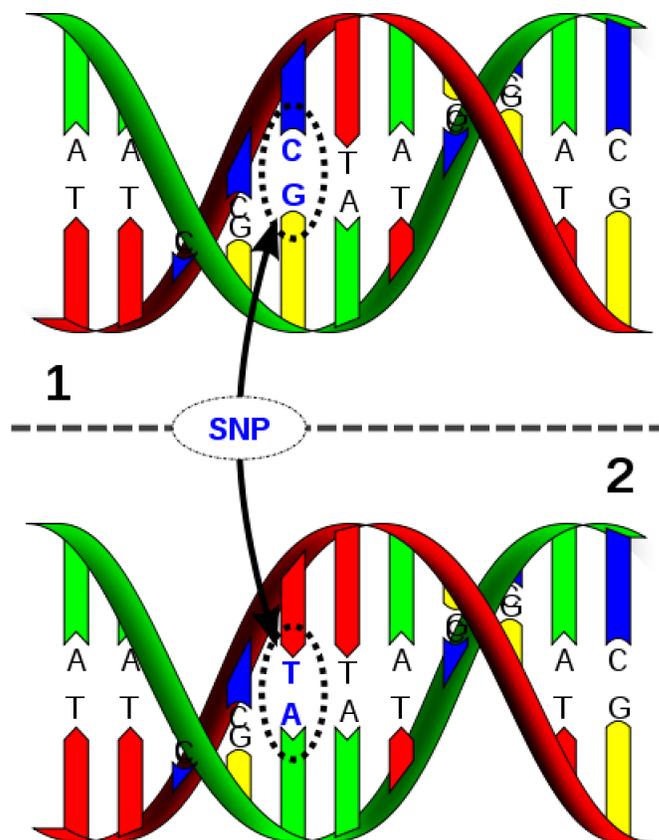


Figure 3: Polymorphisme Nucléotidique Simple (SNP).

Un génotype représente les deux allèles qui composent un SNP particulier chez un individu. Par exemple, pour un SNP donné, les deux allèles possibles pourraient être A et G. Les génotypes possibles dans cette situation sont AA, AG et GG.

2.1.2. expression Quantitative Trait Loci (eQTLs)

Les eQTLs sont des variations génétiques influençant l'expression des gènes. Pour détecter un eQTL, il est nécessaire d'avoir en sa possession les génotypes d'un individu et également les valeurs d'expressions de ce même individu. La détection des eQTLs se fait à l'aide de tests statistiques d'associations. Ces calculs permettent d'identifier des variations génétiques (dans notre étude des SNPs) associées aux niveaux d'expression des gènes. Les eQTLs peuvent être représentés à l'aide de graphique (Figure 4).

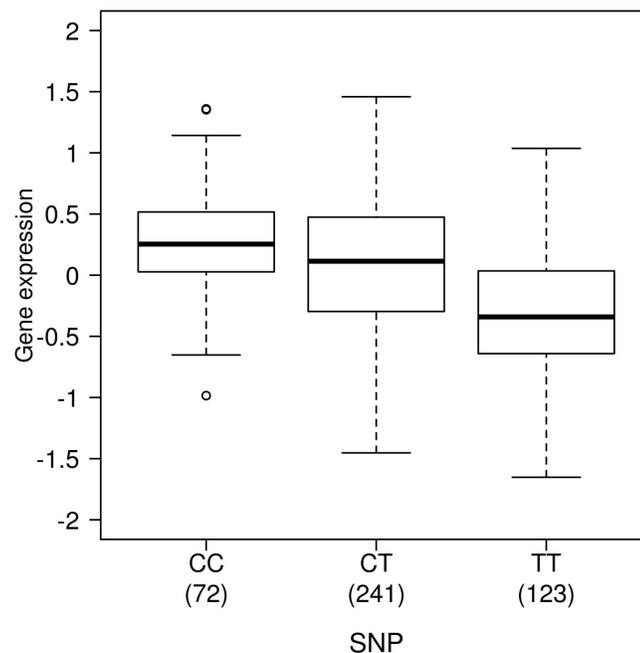


Figure 4: Exemple d'un eQTL. La figure montre un SNP ayant un impact sur l'expression d'un gène. La valeur indiquée sous chaque génotype est le nombre d'individus possédant ce génotype. Dans cet exemple, les individus homozygotes CC ont des niveaux d'expression plus élevés que les homozygotes TT. Les sujets hétérozygotes CT ont des niveaux d'expression intermédiaires.

Les eQTLs se subdivisent en deux catégories : cis-eQTLs et trans-eQTLs. Les cis-eQTLs se caractérisent par une proximité entre le SNP et le gène (Figure 5). Pour un cis-eQTL, le SNP et le gène doivent être à moins de 1 mégabase (Mb) l'un de l'autre sur le même chromosome⁵⁰. Au contraire, le SNP et le gène d'un trans-eQTL doivent être éloignés de plus de 1 Mb ou être situés sur des chromosomes différents.



Figure 5: Représentation schématisée des cis-eQTLs et des trans-eQTLs sur un même chromosome.

2.2. La génétique de la MPOC

Bien que le tabagisme soit le principal facteur dans le développement de la MPOC, ce ne sont pas tous les fumeurs qui vont souffrir de la maladie. Seulement 15 à 20 % d'entre eux vont développer des difficultés respiratoires ce qui sous-entend une susceptibilité génétique aux effets toxiques du tabac¹. Les seuls facteurs génétiques confirmés jusqu'à maintenant menant au développement de la MPOC sont des mutations dans le gène *SERPINA1* résultant en une déficience en alpha-1 antitrypsine. Par contre, au cours des dernières années, plusieurs études génétiques de grandes envergures ont été réalisées. Plusieurs régions chromosomiques et gènes de susceptibilité furent identifiés^{23,24}. Il existe toutefois une grande variabilité dans les résultats obtenus. Cette observation vient appuyer l'idée que la MPOC est une maladie complexe et multifactorielle.

2.2.1. Déficience en alpha-1 antitrypsine

Comme mentionné aux sections précédentes, une déficience en alpha-1 antitrypsine peut mener à la MPOC. La déficience fut identifiée pour la première

fois en 1963 par Carl-Bertil Laurell et Sten Eriksson⁵¹. L'alpha-1 antitrypsine est une anti-protéase qui fait partie de la super-famille des serpinines. Elle a comme fonction d'inhiber un grand éventail de protéases et ainsi protéger les organes contre leurs effets hydrolytiques. La concentration sanguine de l'anti-protéase est de 150 à 350 mg/dL⁵². Le poids moléculaire de la protéine est de 49-kDa ou bien de 57-kDa⁵³.

Pour que l'alpha-1 antitrypsine puisse effectuer son travail correctement, elle doit être présente en quantité suffisante. Ce ne sont pas toutes les mutations qui provoquent des modifications sur les propriétés de l'antitrypsine. Par contre, certaines variations ou délétions vont provoquer un arrêt prématuré de la protéine³³ (Tableau 2).

Tableau 2: Mutations dans le gène SERPINA1*.

Quantité de Protéine / Fréquence Relative	Nom de la Mutation	AA requis	Site de la Mutation (exon)	SNP	AA (Base)	AA (Mutant)
Normale / Commune	M1 (Ala213)			rs6647	Ala213	Ala237
	M1 (Val213)	M1 (Ala213)	III	rs6647	Ala213	Ala237
	M3	M1 (Val213)	V	rs1303	Glu376	Glu400
Normale / Rare	M2	M3	II	rs709932	Arg101	Arg125
	M4	M1 (Val213)	II	rs709932	Arg101	Arg125
	B alhambra	Inconnu	Inconnu	x	x	x
	F	M1 (Val213)	III	rs28929470	Arg223	Arg247
	P saint albans	M1 (Val213)	V	rs143370956	Asp341	Asp365
			III	rs121912714	Asp256	Asp280
	V munich	M1 (Val213)	II	x	Asp2	x
	X	M1 (Val213)	III	x	Glu204	x
	X christchurch	Inconnu	V	rs121912712	Glu363	Glu387
Risque / Commune	Z	M1 (Ala213)	V	rs28929474	Glu242	Glu366
	S	M1 (Val213)	III	rs17580	Glu264	Glu288
Risque / Rare	M heerlen	M1 (Ala213)	V	rs61761869	Pro369	Pro393
	M malton	M2	II	x	Phe52	x
	M mineral springs	M1 (Ala213)	II	rs28931568	Gly67	Gly91
	M procida	M1 (Val213)	II	rs28931569	Leu41	Leu65
	M nichinan	Inconnu	II	x	Phe52	x
			x	rs112030253	Gly148	Gly172
	I	M1 (Val213)	II	x	Arg39	x
	P lowell	M1 (Val213)	III	rs121912714	Asp256	Asp280
	W bethesda	M1 (Ala213)	V	rs1802959	Ala336	Ala360
	Null granite falls	M1 (Ala213)	II	x	Tyr160	x
	Null bellingham	M1 (Val213)	III	x	Lys217	x
	Null matawa	M1 (Val213)	II	rs28929473	Leu353	Leu377
	Null isola di procida	Inconnu	II-V	x	x	x

Null hong kong	M2	IV	x	Leu318	x
Null bolton		V	rs143329723	Pro362	Pho386
		x	rs12233	Pro362	Pho386
Null devon	Inconnu	II	rs11558261	Gly115	Gly139
Null ludwigshafen	M2	II	rs28931572	Ile92	Ile116

*Informations tirées de Crystal, R. G., 1990³³.

2.2.1.1. Diagnostique de la déficience

Il existe trois méthodes différentes pour déterminer si une personne souffre de la déficience : mesurer la quantité de protéine présente dans le sang, caractériser la protéine par focalisation isoélectrique et effectuer un test génétique³⁶.

La quantité normale de protéine dans le sang est de 150 à 350 mg/dL. Une personne souffrant de la déficience aura des niveaux de protéines entre 10 et 80 % par rapport aux niveaux normaux¹². Lorsqu'un médecin soupçonne une déficience, la quantification de la protéine dans le sang est le premier test à effectuer. Cependant, les niveaux d'alpha-1 antitrypsine peuvent varier légèrement lors d'épisodes inflammatoires⁵⁴.

La caractérisation de la protéine se fait en analysant les isoformes de la protéine en laboratoire. Par contre, ce test devient inutile lorsque la protéine n'est pas exprimée¹⁶. De plus, ce test fournit seulement les génotypes présomptifs et a aussi une résolution limitée, c'est à dire qu'il est capable de caractériser seulement un nombre limité de phénotypes protéiques. Ce test exige également que les niveaux sériques de la protéine soient corrélés avec les résultats obtenus, qui sont parfois difficiles à interpréter.

La façon la plus précise d'identifier une déficience est par des tests génétiques. Plusieurs techniques sont disponibles mais souvent limitées à génotyper les deux variations génétiques les plus courantes causant ce phénotype, soit les variations S et Z. Ces deux variations sont responsables de la grande majorité des cas de déficience¹⁶. En plus de ces deux variations, il existe un grand nombre de polymorphismes rares connus depuis plusieurs années qui ne sont pas testés dans la plupart des tests actuels³³ (Tableau 2). De nouvelles

mutations ont également été découvertes encore récemment⁵⁵.

2.2.2. Les études par gènes candidats

Plusieurs études en génétique de la MPOC ont été réalisées à l'aide d'une approche par gènes candidats. Ce genre d'étude cible des gènes dans des voies métaboliques soupçonnées de jouer un rôle dans la maladie. Les voies biologiques étudiées jusqu'à ce jour sont principalement celles de l'inflammation, du système immunitaire, du stress antioxydant, du métabolisme des xénobiotiques et de l'équilibre protéase-antiprotéase²³. Depuis l'arrivée des criblages génomiques par association (GWAS), les études par gènes candidats servent aussi à confirmer ou infirmer les résultats des GWAS. Ce type d'étude a permis d'identifier un grand nombre de gènes associés à divers maladies. Les GWAS seront décrit à la section 2.2.3.

Plusieurs gènes furent identifiés à l'aide d'une approche par gènes candidats, mais les résultats ne sont pas toujours constants d'une étude à l'autre. Plusieurs aspects de ces études peuvent expliquer ces observations : le phénotype utilisé, la population à l'étude, la puissance statistique, une mauvaise interprétation des résultats, les erreurs de génotypage et une mauvaise couverture du gène²³. Les études par gènes candidats ont permis avant la venue des GWAS d'identifier plusieurs gènes et régions chromosomiques possiblement associés à la MPOC et aux fonctions respiratoires (Figure 6).

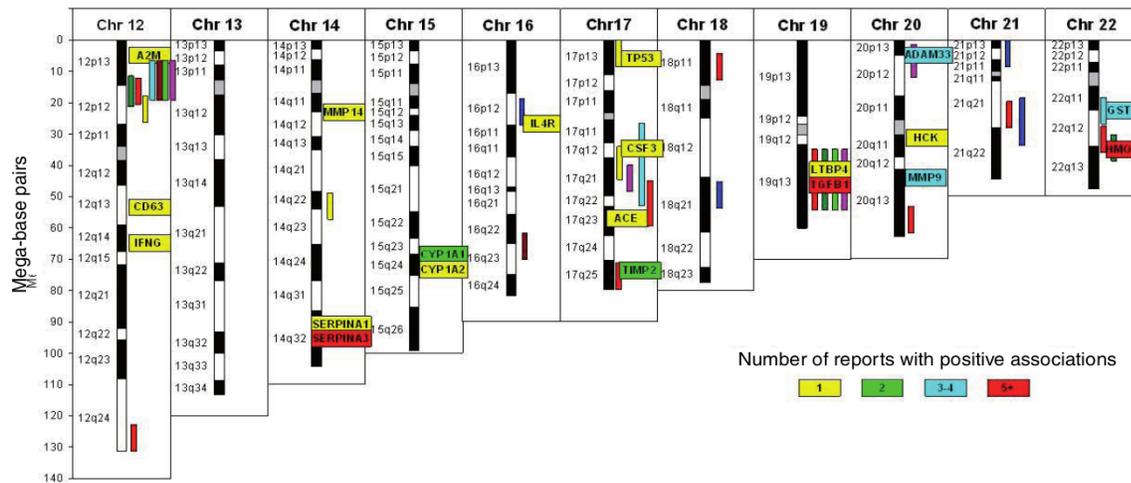
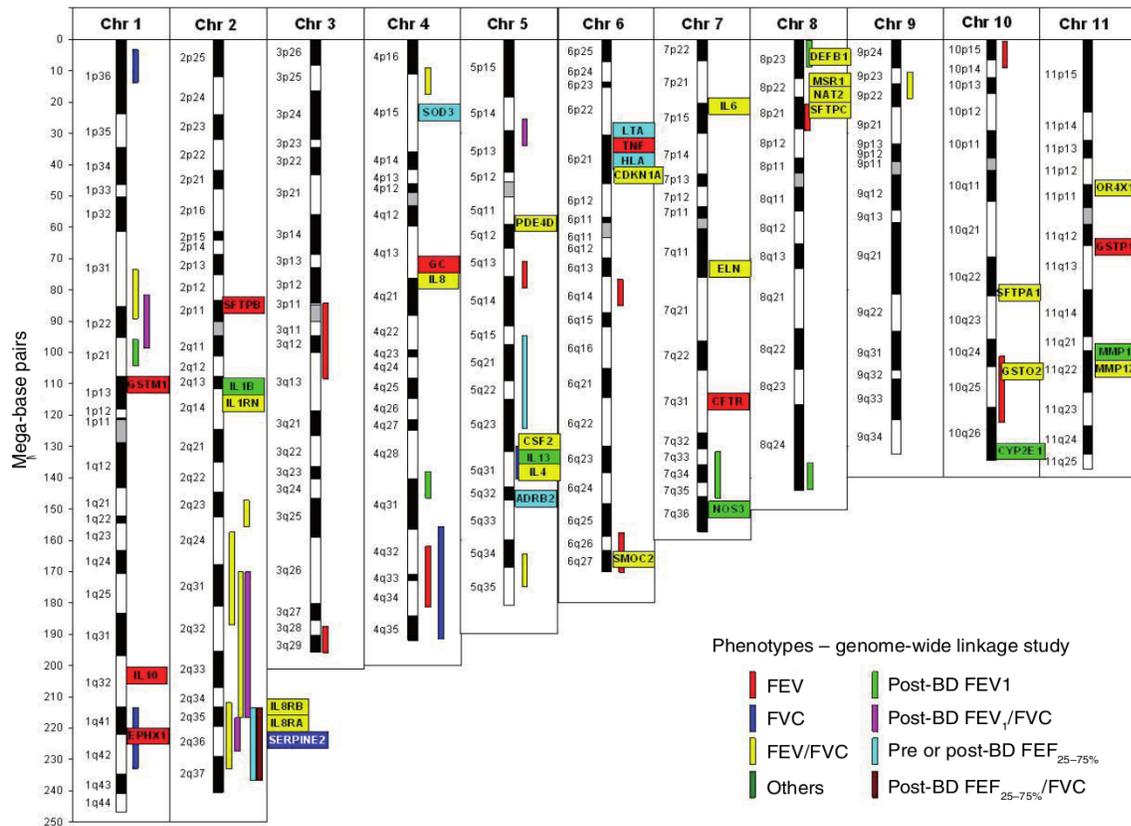


Figure 6 : Gènes candidats de la MPOC. La figure est une représentation graphique de tous les chromosomes humains avec la position des gènes et des loci associés à la MPOC et aux fonctions pulmonaires. Les gènes sont représentés à l'aide de rectangles colorés selon le nombre d'associations positives rapporté dans la littérature scientifique. Le jaune, vert, bleu pâle et rouge indiquent 1, 2, 3 ou 4, et 5 associations ou plus. Le rectangle bleu foncé sur le chromosome 2 représente un gène candidat identifié par position. Les rectangles minces indiquent des loci dérivés des criblages génomiques par liaisons. Ce dernier type d'étude n'est pas traité dans ce mémoire. Figure tirée de Bossé Y., 2009²³.

Le nombre de gènes associés à la MPOC et aux phénotypes reliés à cette maladie est en pleine évolution. En 2009, on comptait 57 gènes qui avaient été associés à la MPOC ou phénotypes reliés à la MPOC²³. Ces gènes sont illustrés à la Figure 6. Une plus récente synthèse de la littérature publiée en 2012 rapporte 144 gènes associés positivement aux mêmes phénotypes dans au moins un article scientifique²⁴. Ces gènes sont illustrés à la Figure 7. Ces chiffres démontrent bien que ce domaine de recherche est en expansion. Il est intéressant de constater que sept gènes ont été associés aux phénotypes de la MPOC dans plus de dix articles scientifiques dont *ADRB2*, *SERPINA1*, *TGFB1*, *TNF*, *GSTM1*, *EPHX1*, et *GSTP1*. Par contre, il faut également souligner que ces gènes n'ont pas été répliqués dans certaines études. Malgré les efforts considérables, nous devons constater que les résultats dérivés de l'approche par gène candidat dans le domaine de la MPOC n'ont pas toujours été concluants.

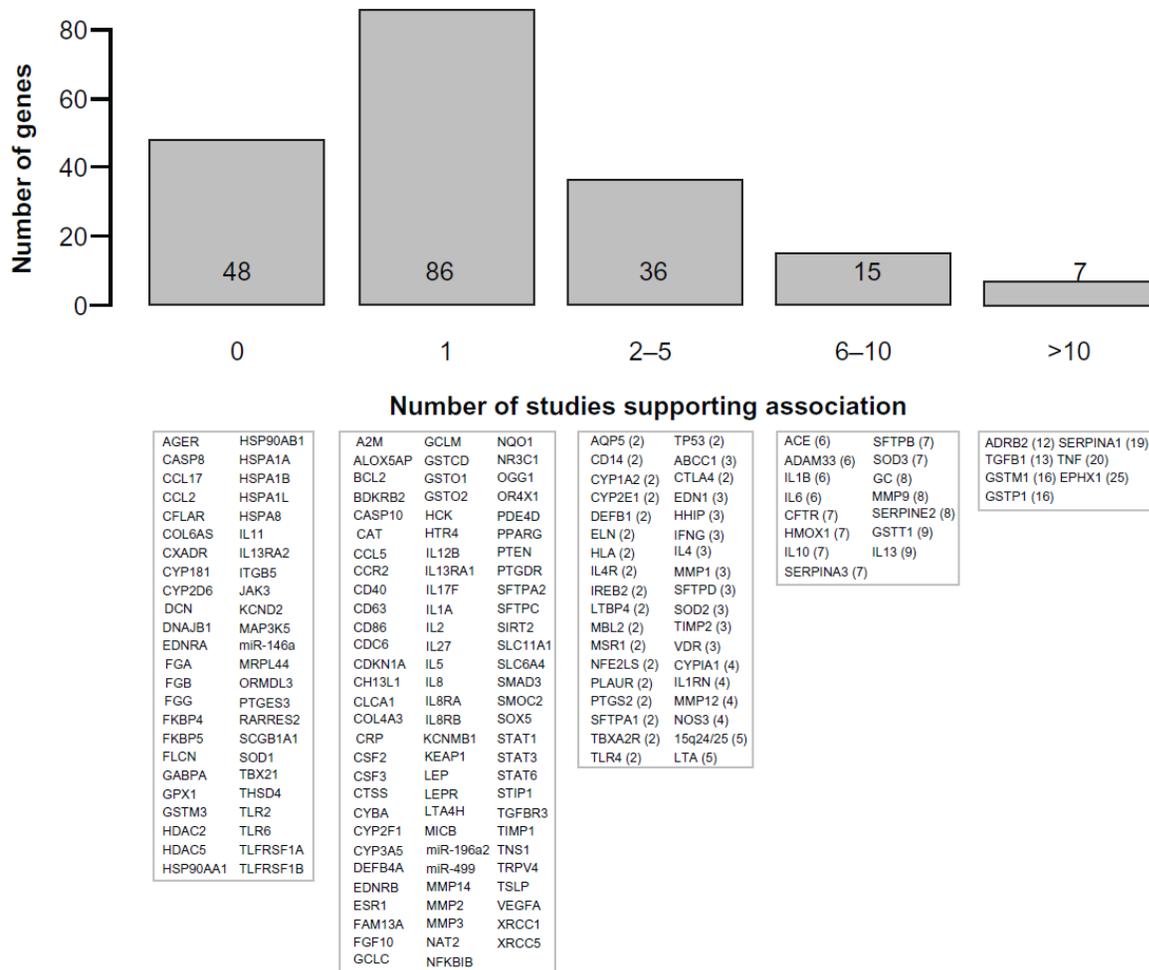


Figure 7: Gènes candidats associés à la MPOC et aux phénotypes liés à cette maladie. La partie supérieure de la figure présente un histogramme du nombre de gènes de susceptibilité à la MPOC (axe des y) sur le nombre d'articles scientifiques (axe des x) ayant démontré une association génétique significative. La partie inférieure de la figure présente les gènes correspondants sous chacune des boîtes. Le nombre de publications qui supporte chaque gène est indiqué entre parenthèses. Figure tiré de Bossé Y., 2012²⁴.

2.2.3. Les criblages génomiques par association (GWAS)

Durant les dernières années, un nombre grandissant de criblages génomiques par association ont été effectués en utilisant différents phénotypes associés à la MPOC⁵²⁻⁶¹. Cette approche consiste à vérifier l'association entre

plusieurs centaines de milliers de polymorphismes répartis sur le génome et un phénotype précis. Les phénotypes étudiés comprennent la MPOC comme trait dichotomique (i.e. cas-témoin), les fonctions respiratoires et l'emphysème. Plusieurs gènes et régions chromosomiques ont été identifiés comme pouvant avoir un impact sur le développement de la MPOC (Tableau 3). Ce genre d'étude a recours à des cohortes de plusieurs milliers d'individus provenant souvent de plusieurs régions différentes. Le principal avantage est que la puissance statistique est beaucoup plus élevée. Par contre, cela ne se fait pas sans risque. La section 2.3 énumère les limitations associées à cette méthode.

Tableau 3: Gènes et régions chromosomiques associées à la MPOC et aux phénotypes reliés par GWAS.

Locus	Gène	Phénotype	Référence
1p36.13	<i>MFAP2</i>	VEMS/CVF	Artigas <i>et al.</i> Nature Genetics. 2011. ⁶⁰
1q41	<i>TGFB2</i>	VEMS/CVF	Artigas <i>et al.</i> Nature Genetics. 2011. ⁶⁰
2q35	<i>TNS1</i>	VEMS	Repapi <i>et al.</i> Nature Genetics. 2010. ⁵⁷
2q36.3	<i>PID1</i>	VEMS/CVF	Hancock <i>et al.</i> Nature Genetics. 2010. ⁵⁶
2q37.3	<i>HDAC4</i>	VEMS/CVF	Artigas <i>et al.</i> Nature Genetics. 2011. ⁶⁰
3p24.2	<i>RARB</i>	VEMS/CVF	Artigas <i>et al.</i> Nature Genetics. 2011. ⁶⁰
3q26.2	<i>MECOM</i>	VEMS	Artigas <i>et al.</i> Nature Genetics. 2011. ⁶⁰
4q22.1	<i>FAM13A</i>	MPOC VEMS/CVF	Cho <i>et al.</i> Nature Genetics. 2010. ⁶⁴ Hancock <i>et al.</i> Nature Genetics. 2010. ⁵⁶
4q24	<i>INTS12</i>	VEMS	Hancock <i>et al.</i> Nature Genetics. 2010. ⁵⁶
4q24	<i>GSTCD</i>	VEMS	Repapi <i>et al.</i> Nature Genetics. 2010. ⁵⁷ Hancock <i>et al.</i> Nature Genetics. 2010. ⁵⁶
4q24	<i>NPNT</i>	VEMS	Hancock <i>et al.</i> Nature Genetics. 2010. ⁵⁶
4q31.21	<i>HHIP</i>	MPOC VEMS/CVF	Pillai <i>et al.</i> PLoS Genetics. 2009. ⁵⁹ Repapi <i>et al.</i> Nature Genetics. 2010. ⁵⁷ Hancock <i>et al.</i> Nature Genetics. 2010. ⁵⁶ Wilk <i>et al.</i> PLoS Genetics. 2009. ⁵⁸
5q15	<i>SPATA9</i>	VEMS/CVF	Artigas <i>et al.</i> Nature Genetics. 2011. ⁶⁰

5q32	<i>HTR4</i>	VEMS VEMS/CVF	Repapi <i>et al.</i> Nature Genetics. 2010. ⁵⁷ Hancock <i>et al.</i> Nature Genetics. 2010. ⁵⁶
5q33.3	<i>ADAM19</i>	VEMS/CVF	Hancock <i>et al.</i> Nature Genetics. 2010. ⁵⁶
6q21	<i>ARMC2</i>	VEMS/CVF	Artigas <i>et al.</i> Nature Genetics. 2011. ⁶⁰
6p21.32	<i>AGER</i>	VEMS/CVF	Repapi <i>et al.</i> Nature Genetics. 2010. ⁵⁷ Hancock <i>et al.</i> Nature Genetics. 2010. ⁵⁶
6p21.2	<i>DAAM2</i>	VEMS/CVF	Repapi <i>et al.</i> Nature Genetics. 2010. ⁵⁷
6p21.32	<i>PPT2</i>	VEMS/CVF	Hancock <i>et al.</i> Nature Genetics. 2010. ⁵⁶
6p21.33	<i>NCR3</i>	VEMS/CVF	Artigas <i>et al.</i> Nature Genetics. 2011. ⁶⁰
6p22.1	<i>ZKSCAN3</i>	VEMS	Artigas <i>et al.</i> Nature Genetics. 2011. ⁶⁰
6q24.1	<i>GPR126</i>	VEMS/CVF	Hancock <i>et al.</i> Nature Genetics. 2010. ⁵⁶
9q22.32	<i>PTCH1</i>	VEMS/CVF	Hancock <i>et al.</i> Nature Genetics. 2010. ⁵⁶
10p13	<i>CDC123</i>	VEMS VEMS/CVF	Artigas <i>et al.</i> Nature Genetics. 2011. ⁶⁰
10q22.2/.3	<i>C10orf11</i>	VEMS	Artigas <i>et al.</i> Nature Genetics. 2011. ⁶⁰
10q25.1	<i>GSTO2</i>	CVF	Wilk <i>et al.</i> BMC Medical Genetics. ⁶¹
12p11.21	<i>BICD1</i>	Emphysème	Kong <i>et al.</i> AJRCCM. 2011. ⁶²
12q13.3	<i>LRP1</i>	VEMS/CVF	Artigas <i>et al.</i> Nature Genetics. 2011. ⁶⁰
12q23.1	<i>CCDC38</i>	VEMS/CVF	Artigas <i>et al.</i> Nature Genetics. 2011. ⁶⁰
15q23	<i>THSD4</i>	VEMS/CVF	Repapi <i>et al.</i> Nature Genetics. 2010. ⁵⁷
15q25.1	<i>AGPHD1</i>	MPOC	Pillai <i>et al.</i> PLoS Genetics. 2009. ⁵⁹
15q25.1	<i>CHRNA3</i>	MPOC	Pillai <i>et al.</i> PLoS Genetics. 2009. ⁵⁹
16q21	<i>MMP15</i>	VEMS/CVF	Artigas <i>et al.</i> Nature Genetics. 2011. ⁶⁰
16q23.1	<i>CFDP1</i>	VEMS/CVF	Artigas <i>et al.</i> Nature Genetics. 2011. ⁶⁰
19q13.2	<i>RAB4B</i>	MPOC	Cho <i>et al.</i> Human Molecular Genetics. 2012. ⁶³
21q22.11	<i>KCNE2</i>	VEMS/CVF	Artigas <i>et al.</i> Nature Genetics. 2011. ⁶⁰

Les GWAS ont été extrêmement utiles pour identifier de nouveaux gènes associés à la MPOC. Par contre, les gènes/loci identifiés par cette approche permettent d'expliquer seulement une faible partie du risque génétique associé à la MPOC^{60,63}. Quatre loci ont été associés directement à la MPOC dans trois GWAS : 4q22.1⁶⁴, 4q31.21⁵⁹, 15q25.1⁵⁹, 19q13.2⁶³. Ces études ont été réalisées chez des patients MPOC de stade modérés à très sévères de descendance européenne. Les tailles d'échantillons étaient entre 2000 à 3500 cas et 1400 à 3400 contrôles. Bien que non négligeable, ce nombre est encore loin des dizaines de milliers de sujets utilisés dans les GWAS sur les fonctions respiratoires^{56,57,60}. De plus, les gènes causals et les variations génétiques fonctionnelles des loci identifiés par GWAS restent en grande majorité inconnus.

Présentement, les études par gènes candidats sont souvent utilisées pour répliquer des associations établies dans d'autres études. Répliquer une association génotype-phénotype demeure toujours le « gold standard » pour les études d'association génétique⁶⁵. Ces réplifications sont essentielles afin d'établir si les nouveaux gènes associés ne sont pas le fruit d'une erreur. De plus, ces études de réplifications permettent de vérifier si les mêmes polymorphismes sont en cause et agissent de la même façon dans d'autres populations. Par exemple, une étude parue en 2012 a vérifié si trois variations génétiques précédemment associées à la MPOC dans des populations non Asiatiques se comportaient de la même façon dans une population provenant d'Asie⁶⁶. Deux SNPs sur trois étaient également associés à la MPOC dans leur population chinoise, mais aucune association n'a été détectée pour le dernier SNP. Cela démontre que même si une variation génétique a été associée à une caractéristique clinique dans une étude, il n'est pas garanti que cette association soit répliquée dans d'autres études.

2.3. Les limitations des études génétiques précédentes

Les deux types d'études précédemment décrites ont plusieurs limitations. Que ce soit au niveau du design expérimental ou bien du choix de la cohorte, plusieurs éléments peuvent influencer les résultats.

Pour les études par gènes candidats, plusieurs limitations ont été énumérées dans la section 2.2.2. Au niveau du phénotype étudié, il faut se rappeler que la MPOC est une maladie hétérogène. Par exemple, les résultats d'une étude sur l'emphysème ne seront pas nécessairement transférables à des études sur la bronchite chronique, les fonctions respiratoires ou la MPOC. La maladie est également sous divisée en plusieurs stades selon la sévérité des symptômes. Il faut également tenir compte de la population à l'étude. Les hommes et les femmes n'ont pas les mêmes risques de développer la maladie. L'âge, l'emploi occupé et le milieu de vie sont tous d'autres facteurs qui peuvent influencer le développement de la maladie et qui peuvent avoir un impact sur les résultats obtenus. La puissance statistique est également très importante dans les études génomiques. Plus le nombre de sujet est élevé, plus les tests statistiques seront sensibles. Des effets plus faibles des polymorphismes pourront être détectés dans de grandes cohortes. Les résultats obtenus peuvent parfois être mal interprétés par les investigateurs. Les erreurs de génotypage peuvent survenir dans ce type d'étude. Une erreur en laboratoire peut mener à la détection d'un faux positif. Une mauvaise couverture du gène est également une source de biais. Puisque les investigateurs doivent faire une sélection de polymorphismes dans leur étude, il est fort possible que les variations causales ne soient pas incluses dans l'expérimentation. Ce problème risque de disparaître dans les années à venir avec l'arrivée du séquençage de nouvelle génération⁶⁷.

Dans le cas des associations par criblages génomiques, des limitations existent toujours et elles sont très semblables à celles des études par gènes candidats. Tout d'abord, avoir de grandes cohortes de patients bien caractérisés au niveau clinique est toujours essentiel. Dans les GWAS regroupant plusieurs populations, une attention particulière doit être portée sur le risque relié à la stratification des populations. Il y a un risque que le background génétique des cohortes soit différent. Pour les cohortes collectées dans plus d'un établissement, la MPOC et les autres phénotypes associés ne sont pas toujours définis de la même façon. La limitation reliée à la couverture d'un gène est aussi pertinente dans les GWAS. Bien que le nombre de variations génétiques soit extrêmement

plus élevé, ce ne sont pas tous les gènes dont la couverture est adéquate. Il est également important de noter que les résultats de GWAS peuvent parfois être difficiles à interpréter. Un grand nombre de SNPs se retrouvent dans les régions intergéniques. Il est difficile de spéculer sur les mécanismes moléculaires menant au développement de la MPOC pour les polymorphismes associés à la maladie localisés dans de telles régions. Des études complémentaires sont nécessaires.

3. Matériels et méthodes

3.1. Sujets de recherche

Les sujets de recherche utilisés dans ces travaux de maîtrise proviennent de trois universités différentes : l'Université Laval, l'Université de la Colombie Britannique, l'Université de Groningen. Tous les patients de l'Université Laval ont été opérés à l'Institut Universitaire de Cardiologie et de Pneumologie de Québec (IUCPQ). La plupart d'entre eux ont subi une opération visant à retirer un cancer du poumon. Des échantillons de tissus cancéreux et sains de poumon étaient prélevés et envoyés par la suite à la Banque de tissus du Réseau en Santé Respiratoire du FRQS au site de l'IUCPQ. Le James Hogg Research Center Biobank du St Paul's Hospital a fournis les échantillons pour l'Université de la Colombie Britannique. Les échantillons pulmonaires de l'Université de Groningen provenaient de la banque de tissus locale du Département de Pathologie. Seulement les tissus sains ont été utilisés dans le cadre de cette étude. Plusieurs données cliniques ont également été amassées. Ces caractéristiques cliniques regroupent environ 100 variables différentes. On y retrouve les variables de base comme l'âge, le sexe, le poids, la taille ainsi que plusieurs paramètres sur l'état de santé des patients. Toutes ces données sont disponibles pour de futurs projets.

3.2. Approches génomiques utilisées

3.2.1. Expression des gènes

Pour déterminer les niveaux d'expression des poumons, nous avons eu recours à 500 spécimens pulmonaires extraits lors de chirurgie. Une fois les ARNm extraits, ils furent hybridés sur des puces à ADN. Plus précisément, nous avons utilisé la puce Affymetrix Human Genome U133 personnalisée par la compagnie Rosetta Inpharmatics. Elle interroge 51,562 sondes (ARNm) simultanément. Les étapes de contrôles de qualité ont réduit ce nombre à 38,858 sondes et à 468, 339 et 363 sujets dans les cohortes de Laval, UBC et Groningen, respectivement.

3.2.2. Génotypage

Les génotypes des patients que nous avons étudiés ont été obtenus à l'aide de puces à ADN. La puce utilisée dans le cadre de cette étude est la Illumina Human 1M-Duo. 1,199,187 SNPs sont interrogés sur cette puce. Les étapes de contrôles de qualité ont exclus 247,919 SNPs des analyses subséquentes. 80, 147 et 113 sujets furent également exclus pour les cohortes de Laval, UBC et Groningen.

3.3. Outils bioinformatiques

3.3.1. Linux

Linux est un système d'exploitation très flexible utilisé dans le domaine de la bio-informatique. Linux est particulièrement avantageux car il est gratuit, rapide et très malléable. Il est utile pour gérer de gros jeux de données. Linux met à la disposition des utilisateurs une vaste gamme de commandes (shell script) pour modifier des fichiers. En plus de son propre langage de programmation, Linux peut également combiner d'autres langages informatiques avec le sien (awk, C, perl, R, etc.). Beaucoup de programmes bio-informatiques fonctionnent sous Linux et s'opèrent habituellement à l'aide de lignes de commande. Linux permet de réaliser des analyses à plus grande vitesse que d'autres systèmes d'exploitation comme Windows.

3.3.2. PLINK

PLINK⁶⁸ est un programme bio-informatique disponible gratuitement sur Internet. Ce programme a été spécialement conçu pour faire des analyses génomiques. Il permet de faire une grande variété de manipulations sur les jeux de données générées par les GWAS. PLINK permet d'accéder à des données génomiques dans une variété de formats. Il permet également de manipuler le fichier pour en extraire les données désirées, par exemple les génotypes pour une variation génétique ou pour un individu particulier. On peut également utiliser PLINK en collaboration avec d'autres programmes.

PLINK est également très pratique pour effectuer des contrôles de qualité. Il permet d'analyser une grande variété de propriétés sur chaque SNP : les allèles, la fréquence allélique, l'équilibre Hardy-Weinberg et le pourcentage d'individus qui ont été génotypés avec succès pour chaque SNP. Il permet également d'obtenir diverses informations sur les individus : pourcentage de SNPs qui ont été génotypé avec succès, vérification du sexe et de liens d'apparenté entre les individus.

Ce programme est également conçu pour faire des tests d'associations génétiques. Il permet de faire des analyses cas-contrôles. Il peut aussi servir à faire des associations avec des variables quantitatives telles les fonctions respiratoires. Des co-variables comme l'âge ou le sexe peuvent également être considérées dans les analyses génétiques.

3.3.3. Affymetrix Power Tools

Affymetrix Power Tools (APT) est un regroupement de programmes informatiques développés par la compagnie Affymetrix. Chaque programme a été optimisé afin d'accomplir une tâche précise dans le délai le plus court possible. APT peut être employé pour tous les types de puces Affymetrix, que ce soit pour le génotypage ou la quantification de l'expression des gènes. APT est disponible sur le site d'Affymetrix :

http://www.affymetrix.com/partners_programs/programs/developer/tools/powertools.affx.

3.3.4. STRUCTURE

STRUCTURE⁶⁹ est un programme permettant d'étudier le background génétique des individus à partir de génotypes. En comparant les génotypes d'un sujet en particulier à celui d'individus de populations bien définies, STRUCTURE nous permet de savoir à quelle population celui-ci appartient. Les populations régulièrement utilisées avec ce programme sont d'origines européennes (CEU), africaines (YRI) et asiatiques (CHB+JPT). Les données de ces populations de référence sont rendues disponibles gratuitement par le projet HapMap⁷⁰.

3.3.5. R

R est un programme informatique de statistique gratuitement disponible. Ce programme est très utilisé pour manipuler des fichiers, effectuer des tests statistiques et générer des graphiques. R est grandement exploité en génomique. Un grand nombre de programmeurs contribuent à son développement par le biais de bibliothèques ayant chacune des fonctions utiles. Ce programme peut être téléchargé gratuitement sur le site « *The Comprehensive R Archive Network* » (CRAN).

3.4. Analyses Statistiques

Des analyses ont été effectuées afin d'identifier des eQTLs dans les poumons. Le test de Wald a été utilisé pour faire ces analyses. Celles-ci ont permis d'identifier des SNPs influençant l'expression des gènes dans des régions chromosomiques sélectionnées. Les valeurs de p ont été obtenues pour toutes les combinaisons de SNPs et de gènes dans les régions d'intérêts. La correction de Bonferroni a été utilisée pour définir le seuil de signification pour chaque locus : $1.39e-06$ pour 4q22 ($0.05/(718 \text{ SNPs} \times 50 \text{ sondes})$), $3.57e-06$ pour 4q31 ($0.05/(412 \text{ SNPs} \times 34 \text{ sondes})$), $7.12e-07$ pour 19q13 ($0.05/(739 \text{ SNPs} \times 95 \text{ sondes})$). Toutes les valeurs de p plus petites que ces seuils respectifs étaient considérées significatives. Les calculs ont été effectués avec le programme PLINK.

4. Objectif et Hypothèses

4.1. Objectif

L'objectif de ce travail est d'identifier des variations génétiques ayant un impact sur l'expression des gènes dans les poumons (eQTLs) dans les régions chromosomiques précédemment associées à la MPOC.

4.2. Hypothèses

Des eQTLs seront détectés dans les régions chromosomiques précédemment associées à la MPOC. Certaines variations génétiques auront un impact sur l'expression des gènes présumés causals dans la littérature. Au contraire, d'autres SNPs associés à la MPOC influenceront des gènes jamais soupçonnés dans cette maladie. Ces nouvelles données viendront compléter et raffiner les résultats des GWAS obtenus sur la MPOC.

Chapitre 2 : Refining susceptibility loci of chronic obstructive pulmonary disease with lung eQTLs

Résumé

Les mécanismes moléculaires menant au développement de la MPOC sont encore inconnus. De récents criblages génomiques par association (GWAS) ont identifié des régions chromosomiques associées à la maladie. Cette étude a pour but d'identifier les variations génétiques et les gènes causaux dans les loci précédemment associés. Le profil d'expression des gènes d'échantillons pulmonaires non tumoraux ainsi que les génotypes pour 1,2 million de SNPs ont été analysés chez 409 sujets. Des «expression quantitative trait loci (eQTLs)» pulmonaire ont été identifiés dans trois régions de susceptibilité génétique identifiées par GWAS : 4q31 (*HHIP*), 4q22 (*FAM13A*), et 19q13 (*RAB4B*, *EGLN2*, *MIA*, *CYP2A6*). Les eQTLs significatifs ont par la suite été répliqués dans deux cohortes indépendantes (n=363 et 339). Les résultats obtenus suggèrent que *HHIP* serait le gène causal dans la région 4q31. Les associations les plus significatives détectées dans les régions 4q22 et 19q13 sont avec les gènes *FAM13A* et *EGLN2*, respectivement.

Refining susceptibility loci of chronic obstructive pulmonary disease with lung eQTLs

Maxime Lamontagne¹, Christian Couture¹, Dirkje S. Postma², Wim Timens³, Don D. Sin^{4,5}, Peter D. Pare^{4,5}, David Nickle⁶, Michel Lavolette¹, and Yohan Bossé^{1,7}

1) Centre de recherche de l'Institut universitaire de cardiologie et de pneumologie de Québec, Québec, Canada;

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

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

4) UBC James Hogg Research Center, Providence Heart + Lung Institute, St. Paul's Hospital, Vancouver, BC, Canada;

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

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

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

Address correspondence to:

Yohan Bossé, Ph.D.

Assistant Professor, Laval University

Institut universitaire de cardiologie et de pneumologie de Québec

Pavillon Marguerite-d'Youville, Y4190

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

Abstract

Chronic obstructive pulmonary disease (COPD) is the fourth cause of chronic morbidity and mortality in the United States and Canada. Recent genome-wide association studies (GWAS) have identified robust susceptibility loci associated with COPD. However, the genetic mechanisms mediating the risk within these loci remain to be found. The goal of this study is to identify causal variants and genes within susceptibility loci associated with COPD. Genome-wide gene expression profiles of 500 non-tumor lung specimens were obtained from patients undergoing lung surgery. Blood-DNA from the same patients were genotyped for 1,2 million SNPs. Following genotyping and gene expression quality control filters, 409 samples were analyzed. Lung expression quantitative trait loci (eQTLs) were identified and overlaid to three COPD susceptibility loci derived from GWAS including 4q31 (*HHIP*), 4q22 (*FAM13A*), and 19q13 (*RAB4B*, *EGLN2*, *MIA*, *CYP2A6*). Significant eQTLs were then replicated in two independent datasets (n=363 and 339). Results suggested that *HHIP* is the causal gene on 4q31. SNPs previously associated with COPD and lung function were regulating the mRNA expression of *HHIP* (rs1828591, rs13118928). An association between the mRNA expression levels of *FAM13A* and rs2045517 was also detected on 4q22, but did not reach statistical significance. On 19q13, the best association between suspected COPD genes and SNPs was *EGLN2*-rs4803369. In summary, this study suggests that lung eQTLs reside in genomic regions previously associated with COPD, which is an important step to understand the molecular mechanisms underlying the COPD susceptibility loci. This study also supports the use of eQTLs in a relevant tissue to prioritize causal genes in GWAS nominated loci.

Introduction

Chronic obstructive pulmonary disease (COPD) is the fourth most common cause of death in the US and is predicted to be the third leading cause of mortality in the world by the year 2030^{1,2}. COPD is a complex disease characterized by airflow obstruction that is not fully reversible³. Cigarette smoking is the most important cause of fast decline in pulmonary function, but not all smokers develop the disease⁴. This observation as well as previous genetic epidemiology studies strongly suggest the implication of a genetic component in the development of COPD⁵. The only well-established genetic risk factors are inherited mutations causing α 1-antitrypsin deficiency⁶. However, these mutations occur only in 1-5% of COPD patients⁷.

The number of susceptibility genes in COPD is expanding rapidly with lists tabulated at 57 genes in 2009⁸ and at 144 genes in 2012⁹. Recent genome-wide association studies (GWAS) have identified four susceptibility loci associated with COPD including 15q25 (*CHRNA3/CHRNA5/IREB2*), 4q31 (*HHIP*), 4q22 (*FAM13A*), and 19q13 (*RAB4B, EGLN2, MIA, CYP2A6*)¹⁰⁻¹². The causal genes and genetic mechanisms mediating the risk within these loci remain to be found.

The goal of this study is to identify lung expression quantitative trait loci (eQTL) within COPD susceptibility loci identified by GWAS. As part of the lung eQTL consortium, we have recently performed a genome-wide search for eQTLs in 1,111 human lung samples¹³. A predefine hypothesis of this consortium was that human lung eQTLs will identify the most informative markers and improve the localization of causal variants/genes in GWAS-nominated COPD loci¹⁴.

Methods

Ethics Statement

All lung tissue samples were obtained in accordance with Institutional Review Board guidelines at the three sites: Laval University (Quebec, Canada), University of British-Columbia (Vancouver, Canada) and Groningen University (Groningen, The Netherlands). All patients provided written informed consent and the study was approved by the ethics committees of the Institut universitaire de cardiologie et de pneumologie de Québec 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>).

Study subjects and lung specimens

Study subjects and lung specimens were described previously^{13,15}. Briefly subjects were from three sites, Laval University, University of British Columbia, and University of Groningen (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 Tissue Bank of the Fonds de recherche du Québec – Santé (FRQS) (www.tissuebank.ca); at Groningen, the lung specimens were provided by the local tissue bank of the Department of Pathology, and at UBC, the lung specimens were provided by the James Hogg Research Center Biobank at St Paul's Hospital. COPD diagnosis and severity were determined according to the GOLD recommendations³. Clinical characteristics of subjects by site are shown in Table 1.

Assays

Genome-wide gene expression and genotyping profiles were obtained using a custom Affymetrix array (see GEO platform 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. Standard quality controls were performed as described previously and only subjects that passed genotyping and expression quality controls were used in this study including 409, 363, and 339 subjects from Laval, Groningen, and UBC, respectively¹³.

COPD susceptibility loci

Lung eQTLs were overlaid to COPD susceptibility loci identified by previous GWAS except for the 15q25-*CHRNA3/CHRNA5/IREB2* locus that we reported previously¹⁶. Three COPD loci were considered including 4q22 (*FAM13A*), 4q31 (*HHIP*) and 19q13 (*RAB4B, EGLN2, MIA, CYP2A6*). SNPs associated with COPD from previous GWAS were tabulated for the three loci (Table 2). SNPs genotyped in the lung eQTL consortium located 1 Mb up and downstream of the most distant associated SNPs in both directions were evaluated. Chromosomes 4q22 (88,875,909-90,886,297), 4q31 (144,480,780-146,506,456) and 19q13 (40,292,404-42,302,706) include 718, 412 and 739 SNPs, respectively. Genes residing in the same regions were tested for *cis*-eQTLs including 14 genes on 4q22 (*SPP1, PKD2, ABCG2, PPM1K, HERC6, HERC5, PIGY, HERC3, NAP1L5, FAM13A, TIGD2, GPRIN3, SNCA, MMRN1*), 9 genes on 4q31 (*FREM3, GYPE, GYPB, GYPA, HHIP, ANAPC10, ABCE1, OTUD4, SMAD1*) and 45 genes on 19q13 (*DYRK1B, FBL, FCGBP, PSMC4, ZNF546, ZNF780B, ZNF780A, MAP3K10, TTC9B, CNTD2, AKT2, C19orf47, PLD3, PRX, SERTAD1, SERTAD3, BLVRB, SPTBN4, SHKBP1, LTBP4, NUMBL, ADCK4, ITPKC, C19orf54, SNRPA, EGLN2, CYP2G1P, CYP2B6, CYP2A13, CYP2S1, AXL, HNRNPUL1, TGFB1, B9D2, TMEM91, EXOSC5, BCKDHA, B3GNT8, ATP5SL, LOC100505495, CEACAM21, CEACAM4, CEACAM7, CEACAM5, CEACAM6*).

Statistical analysis

Lung eQTLs were identified using a different model compared to our previous genome-wide lung eQTL mapping study¹³. Expression data were adjusted for age, sex, and smoking status using robust residuals obtained with the `rlm` function in the R statistical package MASS. Residuals 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 PLINK¹⁷ (version 1.07). Association tests were performed with the “`assoc`” command and the Wald test asymptotic p-values were used for further analysis. Each possible combination of SNPs and genes were tested in the three COPD susceptibility loci in the Laval dataset. Significant eQTLs were those passing Bonferroni correction considering the number of SNPs and genes tested at each loci. P value thresholds were set at 1.4×10^{-6} for 4q22 (0.05/(718 SNPs x 50 probesets), 3.6×10^{-6} for 4q31 (0.05/(412 SNPs x 34 probesets) and 7.1×10^{-7} for 19q13 (0.05/(739 SNPs x 95 probesets). Significant eQTLs in Laval were then validated in the UBC and Groningen datasets. P values lowered than 0.05 were considered significant in the replication sets.

Results

Lung eQTLs in the 4q22 locus

718 SNPs and 50 probesets covering 14 genes were located in the defined region on chromosome 4q22. 66 eQTLs were detected in the Laval set (Figure 1, Table S1). These 66 eQTLs consisted of 50 unique SNPs, 8 probesets and 4 genes (*GPRIN3*, *MMRN1*, *PPM1K*, *SNCA*). Significant linkage disequilibrium (LD) was observed among the 50 SNPs (Figure S1). 51 out of 66 eQTLs replicate in both replication cohorts ($P < 0.05$). The strongest association detected in all cohorts was rs17013978 with *PPM1K* (Figure 2). The expression levels of the gene decrease with the number of T allele. In the three cohorts, this SNP explained 50.2 to 57.1% of the gene expression variance of *PPM1K*. The orientation of the effect was the same in the three cohorts. None of the SNPs previously associated with COPD on 4q22 (Table 2) were genotyped in our eQTL dataset, but five of them were in LD ($R^2 > 0.5$) (Figure 3). These five SNPs were not significantly affecting the expression of genes at that locus. However, three polymorphisms in complete LD (rs2045517, rs2869967, rs2869966) and in modest LD ($r^2 = 0.53-0.69$) with COPD SNPs were associated with the expression levels of *FAM13A* ($p = 4.1 \times 10^{-5}$). Although not significant after correction for multiple testing in the Laval dataset, we evaluated this eQTL in the replication sets. The *FAM13A*-rs2045517 eQTL was significant in UBC, but not in Groningen (Figure 4).

Lung eQTLs in the 4q31 locus

412 SNPs and 34 probesets interrogating 9 unique genes were tested around previously associated SNPs on chromosome 4q31. Significant eQTLs are shown in Figure 5 and Table S2. 53 unique SNPs, 4 probesets and 3 genes (*OTUD4*, *HHIP*, *BC029578*) were involved in significant eQTLs. Only eQTLs associated with *BC029578* were replicated in the two replication sets. eQTL-SNPs on chromosome 4q31 are subdivided in two strong LD block (Figure S2). The strongest eQTL in Laval, validated in both replication sets, was rs7667092 with *BC029578* (Figure 6). The expression levels of the gene increase with the number of T allele in all cohorts. In the three cohorts, this SNP explained 7.6 to 12.5% of

the gene expression variance of *BC029578*. However, this polymorphism was not in LD with SNPs previously associated with COPD. Two SNPs (rs1828591, rs13118928) previously associated with COPD were found to affect the expression of *HHIP*. Rs1828591 was the most significant SNPs associated with *HHIP* in the Laval set. This eQTL was replicated in UBC, but not in Groningen (Figure 7). The G allele was associated with lower expression of *HHIP* in the Laval and UBC set. The same pattern was observed in the Groningen set, but the association was not significant.

Lung eQTLs in the 19q13 locus

On 19q13, 739 SNPs and 95 probesets covering 45 different genes were tested. The expression levels of *RAB4B*, *MIA*, *CYP2A6* were not available in our dataset. 191 eQTLs were detected (Figure 8 and Table S3). 155 SNPs were regulating 9 probesets located on 8 genes (*AXL*, *B3GNT8*, *CEACAM21*, *CEACAM4*, *LOC100505495*, *NUMBL*, *SERTAD3*, *ZNF780A*). 146 eQTLs were validated in both replication cohorts. SNPs associated with gene expression are distributed across four LD blocks (Figure S3). 16 SNPs were associated with the expression levels of *CEACAM21* and *LOC100505495*, and 2 others SNPs were associated with *CEACAM21* and *CEACAM4*. The eQTLs on 19q13 were mainly located in two discrete foci far on both sides of the COPD susceptibility locus *RAB4B/EGLN2/MIA/CYP2A6* (Figure 8). These eQTL-SNPs were not in LD with the COPD SNPs rs7937 and rs2604894. The latter two SNPs are in strong LD ($r^2 = 0.82$) and rs7937 was genotyped in our lung eQTL dataset. Rs7937 was not associated with expression of genes located in this predefined 19q13 locus. The gene mostly regulated by rs7937 was *NUMBL* ($p = 0.0187$). The strongest SNP regulating the expression of suspected COPD genes on 19q33 was rs4803369 modulating the expression of *EGLN2* ($p = 8.9 \times 10^{-7}$). Rs4803369 is located at 13,274 bp away from rs7937 and is in modest LD ($r^2 = 0.33$) with the latter. The eQTL results for rs4803369-*EGLN2* for the three cohorts are illustrated in Figure 9. Although not significant in the Laval dataset, this eQTL was consistently replicated.

Discussion

The goal of this study was to identify causal variants and genes within susceptibility loci associated with COPD. GWAS have identified four loci associated with this disease¹⁰⁻¹². However, the genetic mechanisms mediating the risk within these loci cannot be fully revealed by GWAS. In this study, we used genotypes and expression values of a large number of lung samples derived from three independent populations to identify eQTLs. Our analyses were centered on three loci previously associated with COPD: 4q22 (*FAM13A*), 4q31 (*HHIP*), 19q13 (*RAB4B*, *EGLN2*, *MIA*, *CYP2A6*). We identified genetic variants influencing gene expression at each locus.

The first study to identify an association between the 4q22 locus and COPD was published in 2010¹⁸. Three other studies confirmed an association between this locus and COPD/lung function^{11,12,19}. In this study, we found 66 eQTLs on 4q22 in the original dataset and 51 of them were replicated in both replication sets. The majority of the SNPs were located in intron (n=23) and intergenic region (n=22). Other SNPs were in the 3' UTR (n=3) and upstream gene variant (n=2). Only one missense SNP was found to regulate the expression of *MMRN1*. The three best hits were with the *PPM1K* gene, which were all validated in the replication sets. This gene encodes a member of the PPM family of Mn²⁺/Mg²⁺-dependent protein phosphatases. *PPM1K* has never been associated with lung diseases. SNPs on 4q22 also regulated the expression of *SNCA*. This gene has been associated with accelerated FEV₁/FVC decline²⁰. It is a member of the synuclein family abundantly expressed in the brain. Near *SNCA*, *MMRN1* was also regulated by SNPs in our dataset. *MMRN1* coded for a soluble protein found in platelets and in the endothelium of blood vessels. Interestingly, some SNPs regulating *MMRN1* are located in *SNCA*. Finally, *GPRIN3* was also regulated by SNPs. This gene is located 187 kb from *FAM13A*, which was previously associated with COPD and lung function. Finally, no SNP previously associated with COPD were available in our dataset. However, some of our polymorphisms were in LD with the diseases associated SNPs. The latter were best associated with the expression of *FAM13A*.

Although not significant, this eQTL was validated in one replication set. Accordingly, our results support *FAM13A* as the most likely causal gene on 4q22.

The 4q31 locus was first associated with COPD and lung function in two studies in 2009^{10,21}, and then replicated in four other GWAS^{11,12,18,22}. In our discovery set, 53 SNPs, 4 probesets covering 3 genes were significant. Interestingly, some of the eQTL-SNPs were previously associated with COPD (rs1828591^{10,23}, rs13118928^{10,23}) and lung function (rs1828591²¹, rs7655625²¹, rs1980057^{21,18}). SNPs were mainly located in intergenic region (n=49). Others were in intron (n=3), coding-synonymous (n=1) and 3'UTR (n=1). The only eQTLs replicating in the two replication set are those with the *BC029578* transcript. This transcript is located between the *GYPE* and *GYPB* genes. SNPs regulating *BC029578* are distributed across a 400 kb region. Further studies are needed to understand the function of *BC029578*. eQTLs were also associated with *HHIP*, a member of the hedgehog-interacting protein family. This gene was associated with COPD in the three GWAS on COPD¹⁰⁻¹². Significant eQTLs in this gene only replicated in UBC, but the same pattern were also observed in the Goningen set. These results supported that *HHIP* is the most likely causal gene in the region.

Many genes are present in the 19q13 locus. This locus was recently associated with COPD and so far no replication study was performed¹². 191 eQTLs were detected in our original set and 146 of them were validated in the replication sets. Eight genes were regulated by SNPs in the Laval dataset, which were all validated in replication sets. Some SNPs were previously associated with COPD (rs2302188²⁴, rs4803481²⁴, rs1800469^{25,26}) and lung function (rs2241718²⁵, rs6957²⁵). Interestingly, *CEACAM21* was associated with COPD susceptibility in a sputum eQTLs study on COPD patients²⁴. To the best of our knowledge, no studies supported the contribution of *AXL*, *NUMBL*, *SERTAD3*, *B3GNT8*, *CEACAM4*, *LOC100505495* or *ZNF780A* in COPD or related phenotypes. Rs7937, a SNP identified in previous GWAS, was genotyped in our datasets, but no association was detected between this SNP and the expression level of any gene. The strongest association with a suspected COPD gene is *EGLN2*-rs4803369. This

gene is involved in regulating hypoxia tolerance and apoptosis in cardiac and skeletal muscle. These results supported, not convincingly, that *EGLN2* is the potential causal COPD gene on 19q13.

In conclusion, we used a large collection of human lung specimens to identify SNPs that regulated gene expression in three COPD susceptibility loci derived from GWAS. Strong lung eQTLs were detected in the three COPD loci. However, the eQTL-SNPs were not necessarily in LD with SNPs associated with COPD. On 4q22, SNPs associated with COPD near the *FAM13A* gene were indirectly (in LD) associated with the mRNA expression levels of *FAM13A*. On 4q31, the suspected candidate in this region, *HHIP*, were found to be regulated by SNPs associated with COPD and lung function. On 19q13, SNPs associated with COPD were most strongly associated with the expression level of *EGLN2*. Further functional studies will be needed to verify the contribution of susceptibility genes in COPD. This study is an important step to understand the underlying biology of the COPD susceptibility loci. It also shows the potential of eQTLs in a relevant tissue to leverage the results of previous GWAS and extend their functional meaning to gene expression.

Acknowledgments

The authors would like to thank Christine Racine and Sabrina Biardel at the Respiratory Health Network Tissue Bank of the FRQS for their valuable assistance. We also acknowledge the staff at the Consortium Laval, Université du Québec, McGill and Eastern Quebec (CLUMÉQ) for IT support with the high performance computer clusters. Y. Bossé is a research scholar from the Heart and Stroke Foundation of Canada. This study was partly 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, and the Respiratory Health Network of the FRQS.

References

1. Miniño, A. M., Heron, M. P., Murphy, S. L. & Kochanek, K. D. Deaths: final data for 2004. *National vital statistics reports from the Centers for Disease Control and Prevention National Center for Health Statistics National Vital Statistics System* **55**, 1–119 (2007).
2. Who, W. H. O. *World Health Statistics 2008*. 212 (2008).at
3. Rabe, K. F. *et al.* Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *American Journal of Respiratory and Critical Care Medicine* **176**, 532–555 (2007).
4. Mannino, D. M., Homa, D. M., Akinbami, L. J., Ford, E. S. & Redd, S. C. Chronic obstructive pulmonary disease surveillance--United States, 1971-2000. *Respiratory Care* **47**, 1184–1199 (2002).
5. Silverman, E. K. Progress in Chronic Obstructive Pulmonary Disease Genetics. *Proceedings of the American Thoracic Society* **3**, 405–408 (2006).
6. Ganrot, P. O., Laurell, C. B. & Eriksson, S. Obstructive lung disease and trypsin inhibitors in alpha-1-antitrypsin deficiency. *Scandinavian journal of clinical and laboratory investigation* **19**, 205–208 (1967).
7. Marciniuk, D., Hernandez, P. & Balter, M. Alpha-1 antitrypsin deficiency targeted testing and augmentation therapy: A Canadian Thoracic Society clinical practice guideline. *Canadian Thoracic* 109–116 (2012).
8. Bossé, Y. Genetics of chronic obstructive pulmonary disease: a succinct review, future avenues and prospective clinical applications. *Pharmacogenomics* **10**, 655–667 (2009).
9. Bossé, Y. Updates on the COPD gene list. *International Journal of COPD* **7**, 607–631 (2012).
10. Pillai, S. G. *et al.* A Genome-Wide Association Study in Chronic Obstructive Pulmonary Disease (COPD): Identification of Two Major Susceptibility Loci. *PLoS Genetics* **5**, 8 (2009).
11. Cho, M. H. *et al.* Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nature Genetics* **42**, 200–202 (2010).
12. Cho, M. H. *et al.* A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Human molecular genetics* **21**, 947–57 (2012).
13. Hao, K. *et al.* Lung eQTLs to Help Reveal the Molecular Underpinnings of Asthma. *PLoS Genetics* **8**, e1003029 (2012).

14. Bosse, Y., Sin, D., Laviolette, M. & Sandford, A. Hypothesis-driven research on genomic data derived from a large scale lung eQTL mapping study. *WebmedCentral* **1**, (2010).
15. Bossé, Y. *et al.* Molecular Signature of Smoking in Human Lung Tissues. *Cancer research* 1–11 (2012).
16. Nguyen, J. D. U. *et al.* Susceptibility loci for lung cancer on chromosome 15q25 are associated with mRNAs levels of CHRNA5 in the lung. Submit
17. Purcell, S. *et al.* PLINK: a toolset for whole genome association and population-based linkage analyses. *Genome Research* **81**, 559–75 (2007).
18. Hancock, D. B. *et al.* Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nature Genetics* **42**, 45–52 (2010).
19. Pillai, S. G. *et al.* Loci identified by genome-wide association studies influence different disease-related phenotypes in chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine* **182**, 1498–1505 (2010).
20. Curjuric, I. *et al.* Different genes interact with particulate matter and tobacco smoke exposure in affecting lung function decline in the general population. *PLoS one* **7**, e40175 (2012).
21. Wilk, J. B. *et al.* A Genome-Wide Association Study of Pulmonary Function Measures in the Framingham Heart Study. *PLoS Genetics* **5**, 8 (2009).
22. Repapi, E. *et al.* Genome-wide association study identifies five loci associated with lung function. *Nature Genetics* **42**, 36–44 (2010).
23. Van Durme, Y. M. T. a *et al.* Hedgehog-interacting protein is a COPD susceptibility gene: the Rotterdam Study. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology* **36**, 89–95 (2010).
24. Qiu, W. *et al.* Genetics of sputum gene expression in chronic obstructive pulmonary disease. *PLoS one* **6**, e24395 (2011).
25. Celedón, J. C. *et al.* The transforming growth factor-beta1 (TGFB1) gene is associated with chronic obstructive pulmonary disease (COPD). *Human molecular genetics* **13**, 1649–56 (2004).
26. Ito, M. *et al.* The Association of Transforming Growth Factor Beta 1 Gene Polymorphisms with the Emphysema Phenotype of COPD in Japanese. *Internal Medicine* **47**, 1387–1394 (2008).

Table 1: Clinical characteristic of patients that passed gene expression and genotyping quality control filters.

	Laval (n = 409)	UBC (n = 339)	GRNG (n = 363)
Gender (%)	229 : 180 (55.9 % male)	182 : 157 (53.7 % male)	193 : 170 (53.2 % male)
Age (years)	63.3 ± 9.9	60.2 ± 14.3	51.5 ± 15.5 [9]
Body Mass Index (kg/m ²)	26.7 ± 5.3	25.6 ± 5.4 [56]	23.2 ± 4.2 [42]
FEV ₁ predicted - pre-BD (%)	80.5 ± 18.9 [16]	78.2 ± 24.4 [77]	60.5 ± 30.0 [194]
FVC predicted – pre-BD (%)	89.8 ± 16.1 [31]	86.9 ± 20.1 [75]	75.0 ± 26.5 [208]
COPD	211 (51.6 %) [34]	115 (33.9 %) [99]	159 (43.8 %) [120]
Stage 1 : Mild	82 (38.9 %)	43 (37.4 %)	20 (12.6 %)
Stage 2 : Moderate	117 (55.4 %)	60 (52.2 %)	38 (23.9 %)
Stage 3 : Severe	11 (5.2 %)	2 (1.7 %)	21 (13.2 %)
Stage 4 : Very Severe	1 (0.5 %)	10 (8.7 %)	69 (43.4 %)
Asthma	15 (3.7 %)	22 (6.5 %)	0 (0.0 %)
Diabetes	41 (10.0 %)	13 (3.8 %)	27 (7.4 %)
Cardiac diseases	120 (29.3 %)	46 (13.6 %)	28 (7.7 %)
Smoking		[52 (15.3 %)]	[21 (5.8 %)]
Smoker	90 (22.0 %)	98 (34.1 %)	57 (16.7 %)
Ex-Smoker	283 (69.2 %)	163 (56.8 %)	185 (54.1 %)
Non-Smoker	36 (8.8 %)	26 (9.1 %)	100 (29.3 %)
Pack-year in ever-smokers	48.5 ± 27.5 [37]	44.7 ± 28.5 [58]	31.2 ± 17.4 [51]

FEV₁ : forced expiratory volume in 1 seconde.

FVC : forced vital capacity.

[] = missing value.

Table 2: SNPs associated with COPD in previous GWAS.

Locus	SNP	SNP position	Study
4q22	rs1964516	89,875,909	Cho et al. 2012. Human Molecular Genetics. ¹²
	rs7671167	89,883,979	Cho et al. 2010. Nature Genetics. ¹¹ Cho et al. 2012. Human Molecular Genetics. ¹²
	rs1903003	89,886,297	Cho et al. 2010. Nature Genetics. ¹¹
4q31	rs1828591	145,480,780	Cho et al. 2010. Nature Genetics. ¹¹ Pillai et al. 2009. PLoS Genetics. ¹⁰
	rs13118928	145,486,389	Cho et al. 2012. Human Molecular Genetics. ¹² Pillai et al. 2009. PLoS Genetics. ¹⁰
	rs13141641	145,506,456	Cho et al. 2012. Human Molecular Genetics. ¹²
19q13	rs2604894	41,292,404	Cho et al. 2012. Human Molecular Genetics. ¹²
	rs7937	41,302,706	Cho et al. 2012. Human Molecular Genetics. ¹²

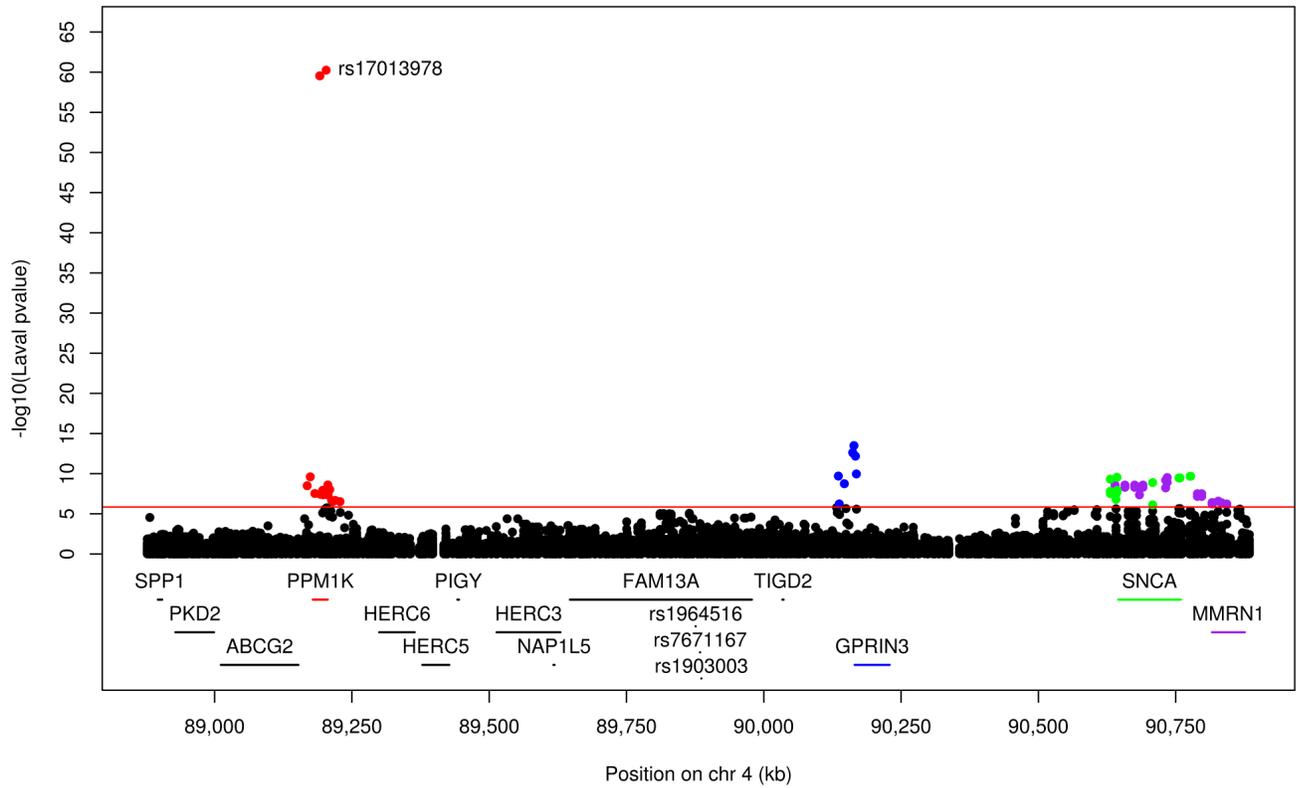


Figure 1: Lung eQTLs on 4q22 in the Laval dataset. Each dot represents an association test between one SNP and one probeset. Only dots above the red line are significant ($p < 1.4 \times 10^{-6}$). Significant SNPs were regulating the expression levels of *PPM1K* in red, *GPRIN3* in blue, *SNCA* in green and *MMRN1* in purple. The SNP with the lower pvalue is indicated. SNPs previously associated with COPD are presented at the bottom.

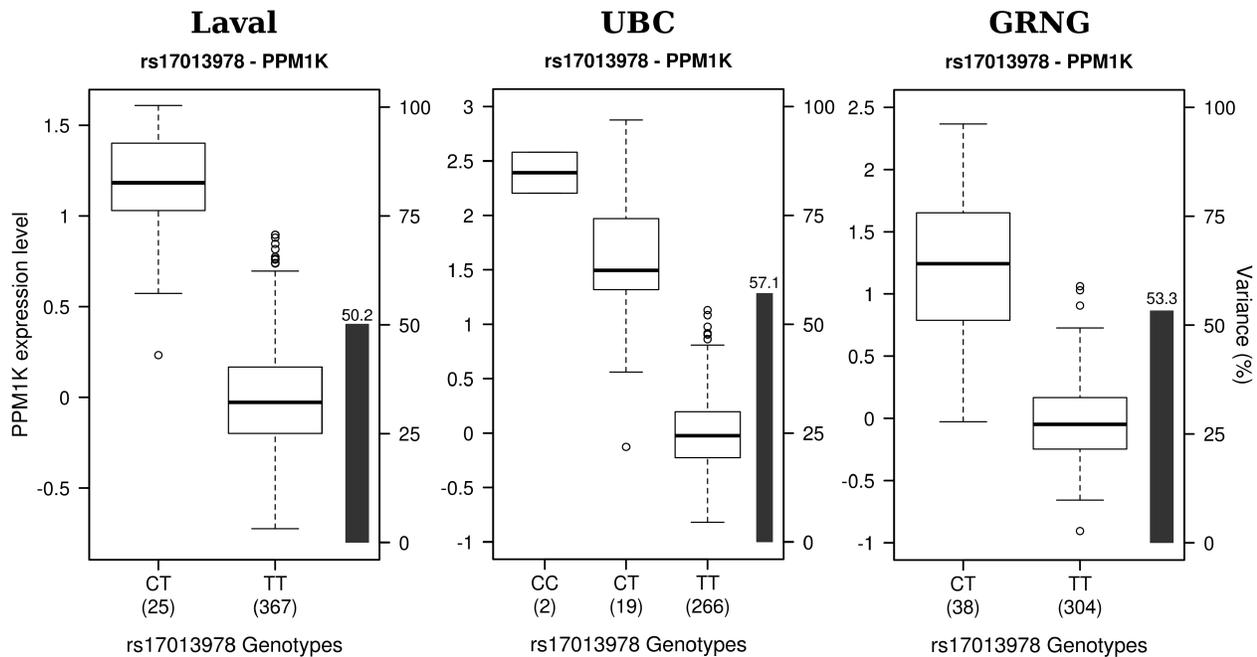


Figure 2: Boxplots of gene expression levels in the lung for *PPM1K* according to genotype groups for SNP rs17013978. The left y-axis shows the mRNA expression levels for *PPM1K*. The x-axis represents the three genotype groups for SNP rs17013978. The right y-axis shows the proportion of the gene expression variance explained by the SNP (black bar). Each panel represents a different cohort: Laval (n=392), UBC (n=287), GRNG (n=342). The eQTL pvalues were 5.6×10^{-61} , 2.8×10^{-51} and 3.8×10^{-55} , respectively.

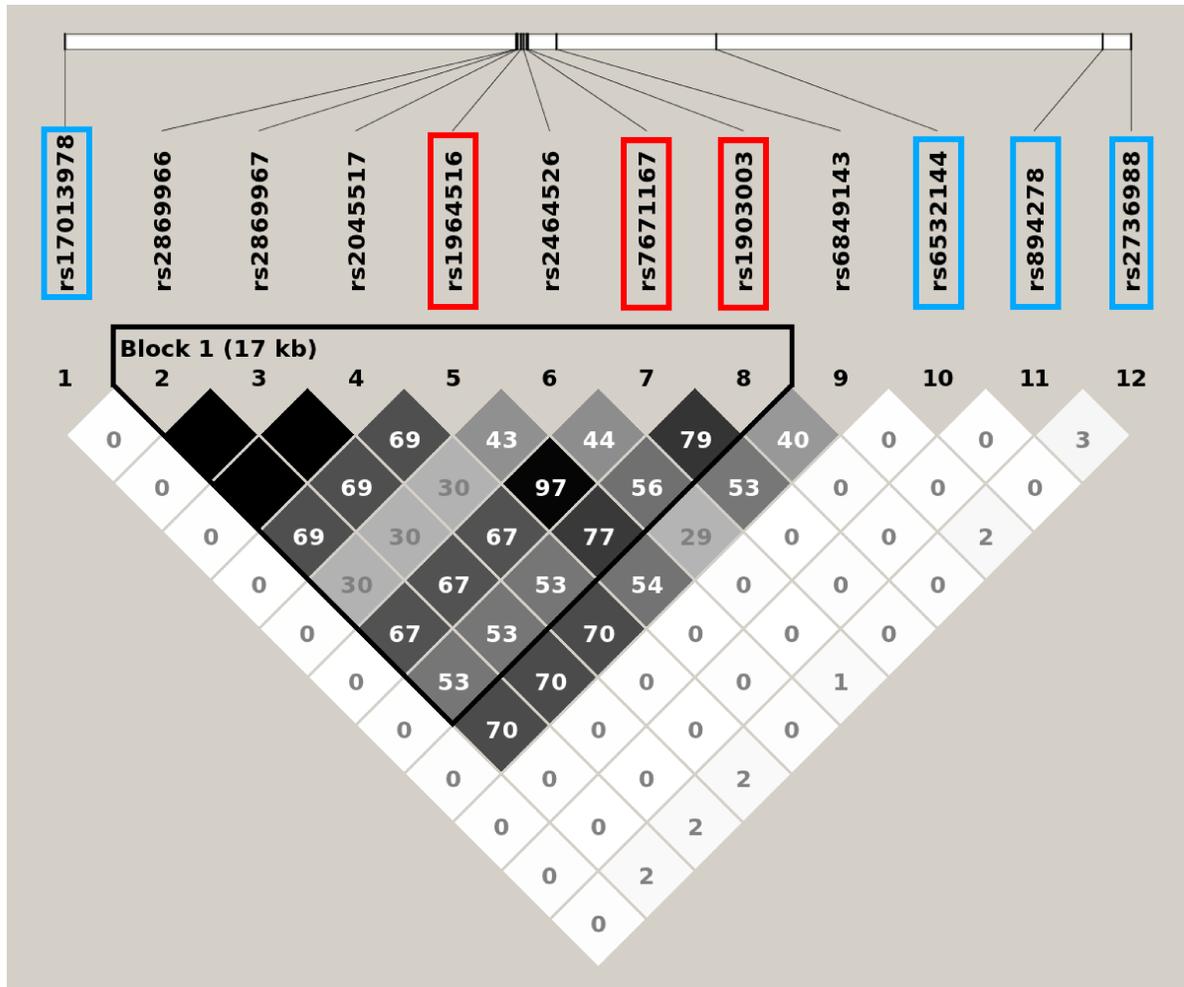


Figure 3: Linkage disequilibrium plot of selected SNPs on the 4q22 locus in the 1000 Genome Project. The white horizontal bar on the upper part of the figure illustrates the location of SNPs on a physical scale. LD values (r^2) are indicated in each box. The color of the squares illustrate the strength of pairwise r^2 values on a black and white scale where black indicates perfect LD ($r^2 = 1$) and white indicates perfect equilibrium ($r^2 = 0$). The genotypes are from the 1000 Genome Project interim phase1 release (2010/11/23). Red rectangles are SNPs previously associated with COPD (Table 2). Blue rectangles are the best eQTL-SNPs for the four regulated genes found on 4q22 (Figure 1). The other SNPs illustrated were genotyped in our study and in LD ($r^2 > 0.5$) with COPD SNPs.

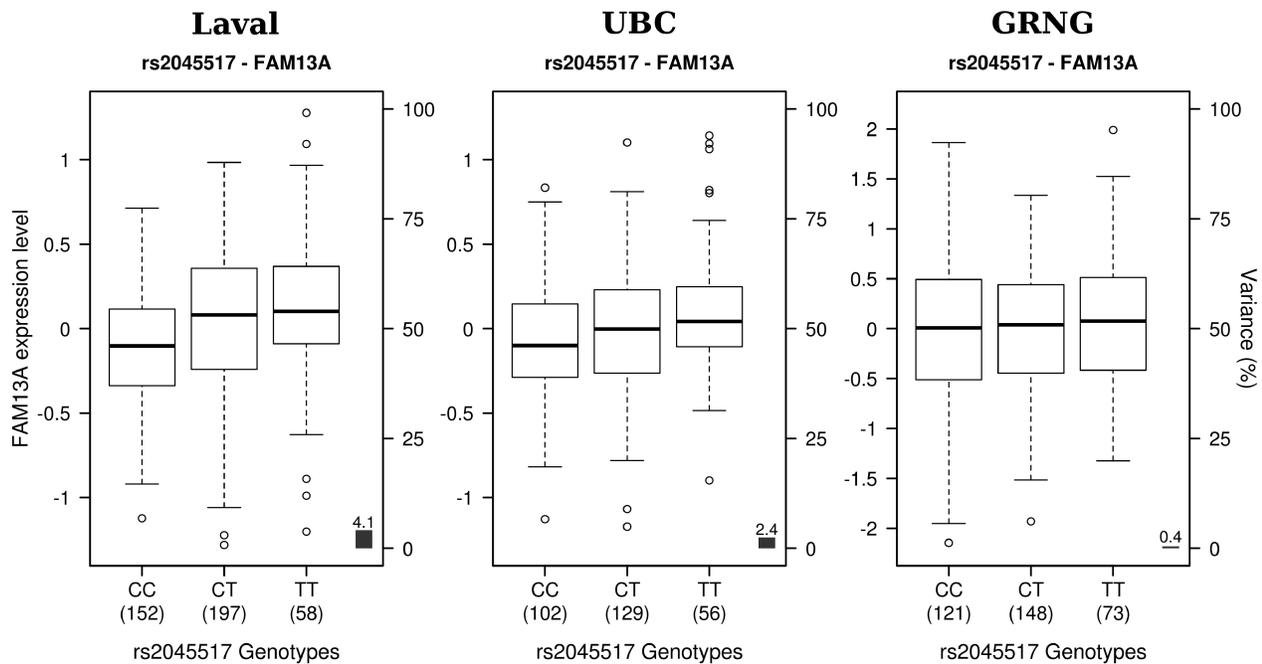


Figure 4: Boxplots of gene expression levels in the lung for *FAM13A* according to genotype groups for SNP rs2045517. The left y-axis shows the mRNA expression levels for *FAM13A*. The x-axis represents the three genotype groups for SNP rs2045517. The right y-axis shows the proportion of the gene expression variance explained by the SNP (black bar). Each panel represents a different cohort: Laval (n=407), UBC (n=287), GRNG (n=342). The eQTL pvalues were 4.1×10^{-5} , 0.009 and 0.218 , respectively.

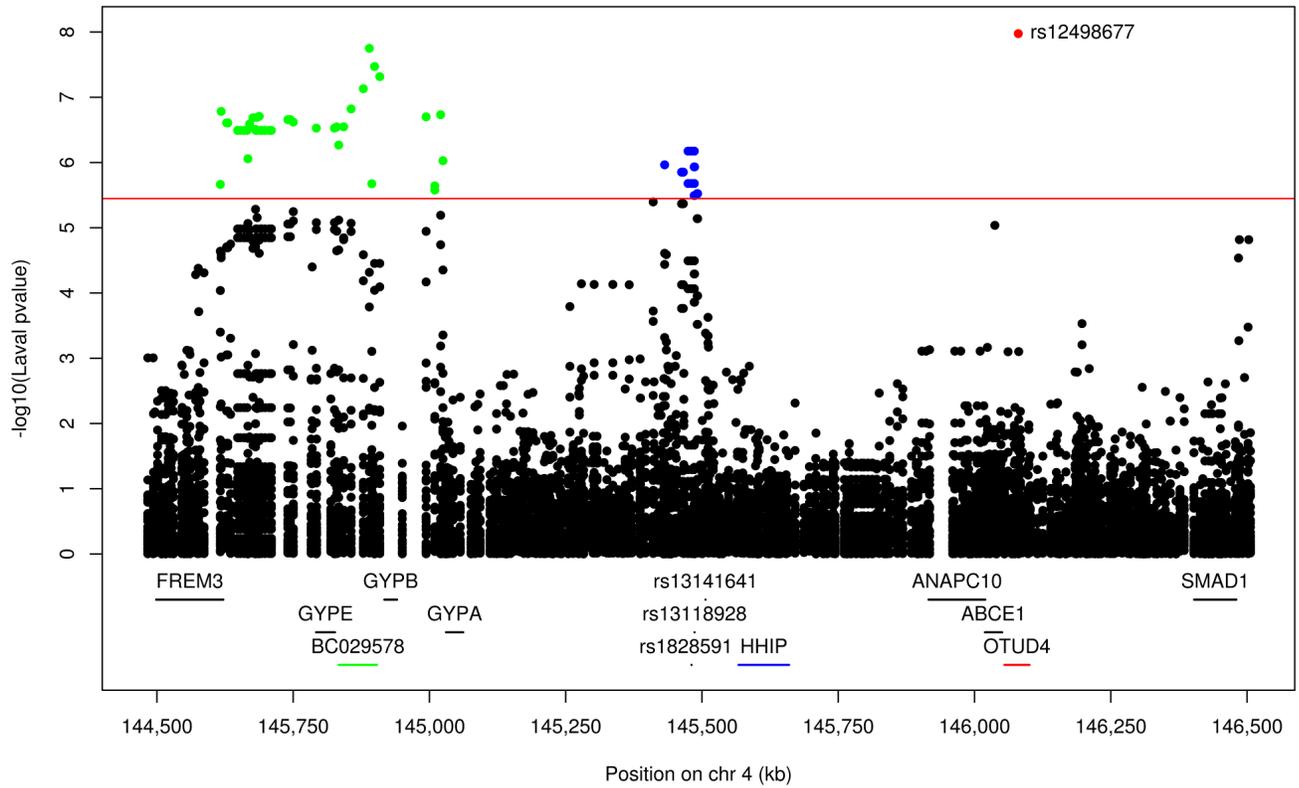


Figure 5: Lung eQTLs on 4q31 in the Laval dataset. Each dot represents an association test between one SNP and one probeset. Only dots above the red line are significant ($p < 3.6 \times 10^{-6}$). Significant SNPs were regulating the expression levels of *HHIP* in blue, *OTUD4* in red, and *BC029578* in green. The SNP with the lower pvalue is indicated. SNPs previously associated with COPD are presented at the bottom.

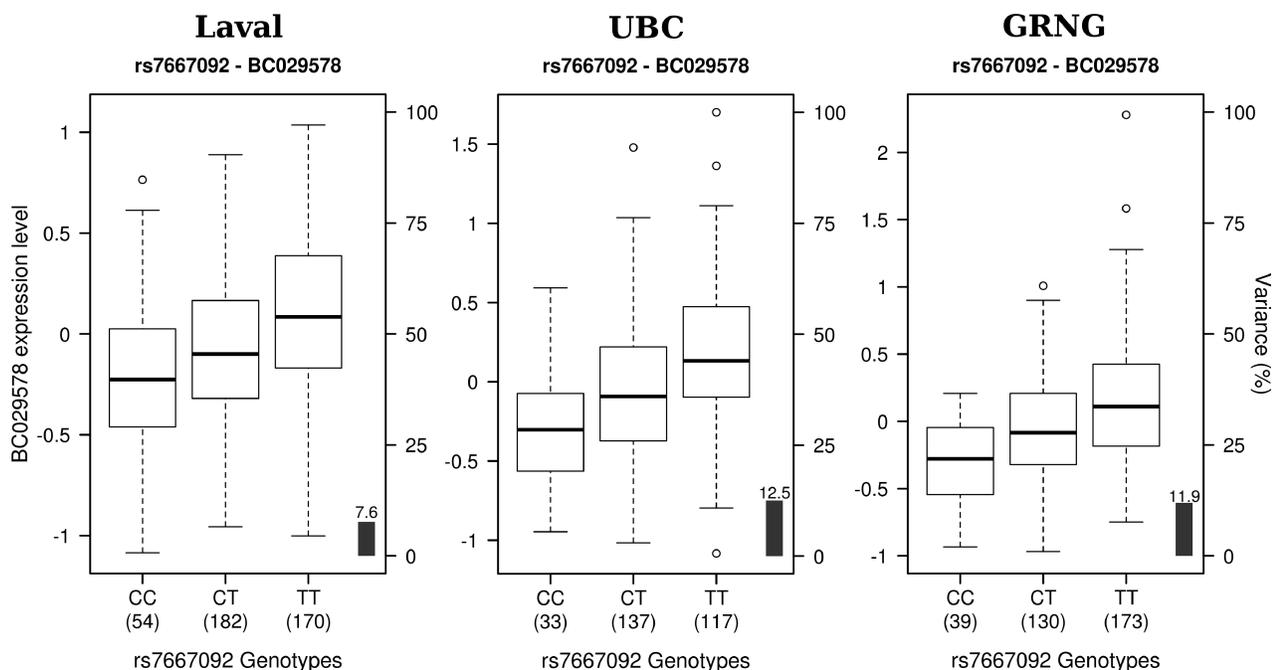


Figure 6: Boxplots of gene expression levels in the lung for *BC029578* according to genotype groups for SNP rs7667092. The left y-axis shows the mRNA expression levels for *BC029578*. The x-axis represents the three genotype groups for SNP rs7667092. The right y-axis shows the proportion of the gene expression variance explained by the SNP (black bar). Each panel represents a different cohort: Laval (n=406), UBC (n=287), GRNG (n=342). The eQTL p-values were 1.8×10^{-8} , 7.3×10^{-10} and 6.0×10^{-11} , respectively.

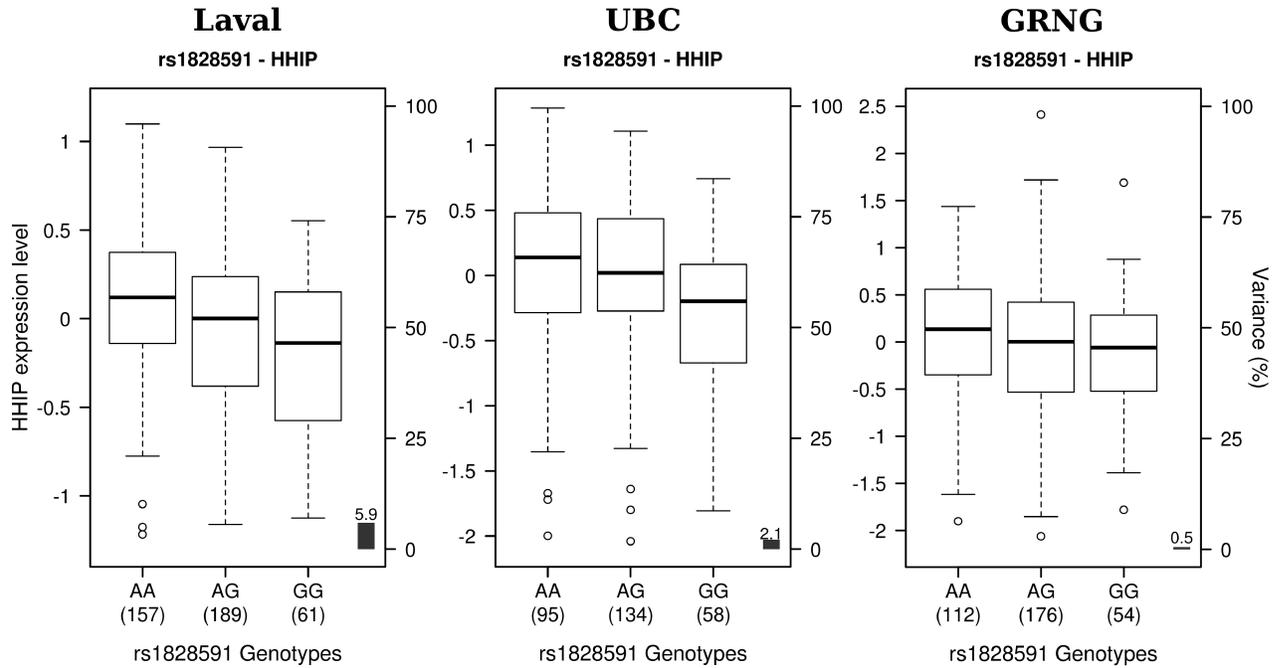


Figure 7: Boxplots of gene expression levels in the lung for *HHIP* according to genotype groups for SNP rs1828591. The left y-axis shows the mRNA expression levels for *HHIP*. The x-axis represents the three genotype groups for SNP rs1828591. The right y-axis shows the proportion of the gene expression variance explained by the SNP (black bar). Each panel represents a different cohort: Laval (n=407), UBC (n=287), GRNG (n=342). The eQTL pvalues were 6.7×10^{-7} , 0.013 and 0.214, respectively.

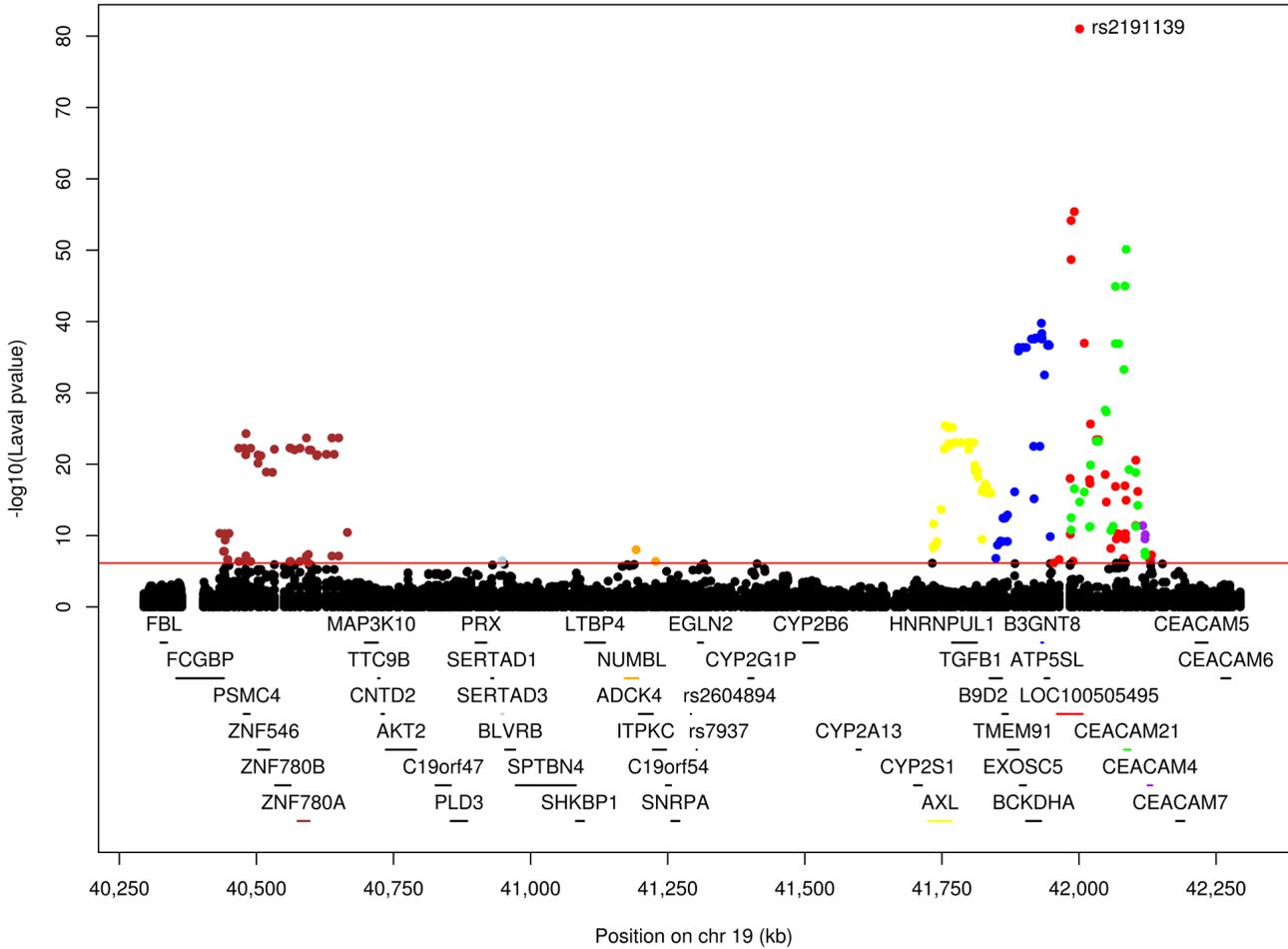


Figure 8: Lung eQTLs on 19q13 in the Laval dataset. Each dot represents an association test between one SNP and one probeset. Only dots above the red line are significant ($p < 7.1 \times 10^{-7}$). Significant SNPs were regulating the expression levels of *ZNF780A* in brown, *SERTAD3* in light blue, *NUMBL* in orange, *AXL* in yellow, *B3GNT8* in blue, *LOC100505495* in red, *CEACAM21* in green, and *CEACAM4* in purple. The SNP with the lower pvalue is indicated. SNPs previously associated with COPD are illustrated at the bottom.

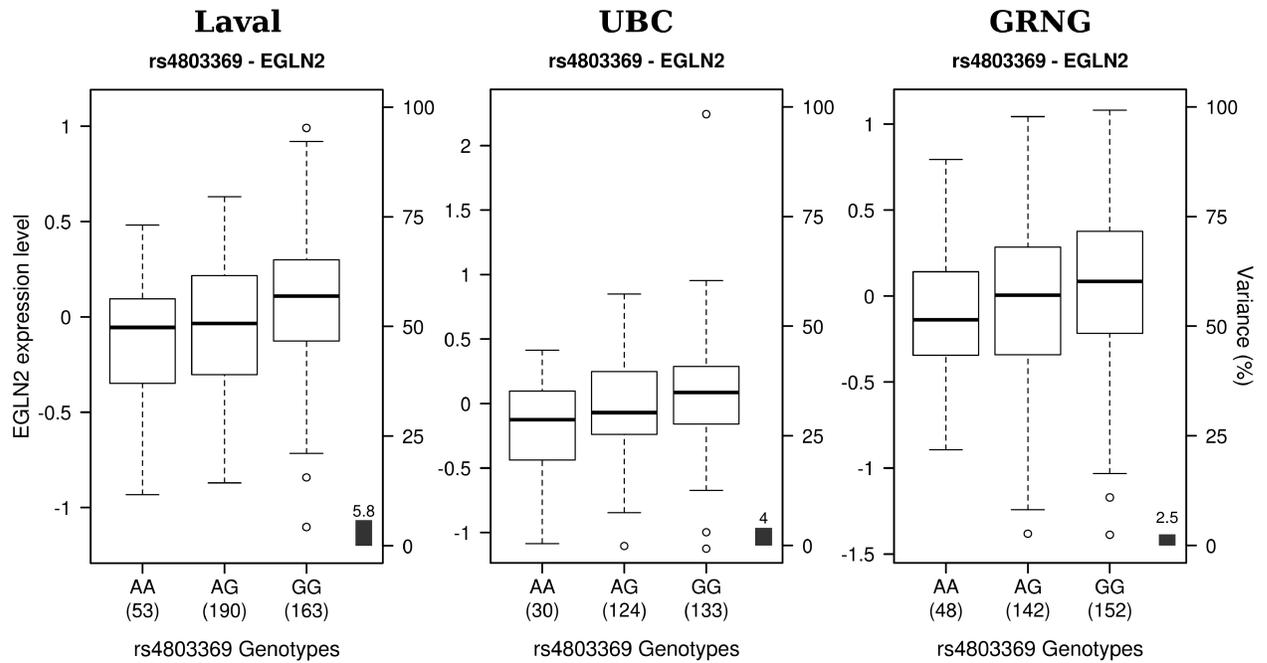


Figure 9: Boxplots of gene expression levels in the lung for *EGLN2* according to genotype groups for SNP rs4803369. The left y-axis shows the mRNA expression levels for *EGLN2*. The x-axis represents the three genotype groups for SNP rs4803369. The right y-axis shows the proportion of the gene expression variance explained by the SNP (black bar). Each panel represents a different cohort: Laval (n=406), UBC (n=287), GRNG (n=342). The eQTL pvalues were 8.9×10^{-7} , 0.001 and 0.003, respectively.

Table S1: Significant eQTLs at the 4q22 locus in the Laval dataset and replication in UBC and Groningen.

SNP	SNP Position	SNP Function	Regulated Gene	P Laval	P UBC	P GRNG
rs17013978	89203197	intron	PPM1K	5.636e-61	2.79e-51	3.77e-55
rs7682161	89191346	intron	PPM1K	2.876e-60	2.79e-51	3.77e-55
rs7658312	89191741	intron	PPM1K	2.876e-60	2.79e-51	3.77e-55
rs6532144	90164136	unknown	GPRIN3	3.167e-14	1.629e-06	0.003492
rs7660707	90161863	unknown	GPRIN3	2.319e-13	1.734e-09	0.0006004
rs17015234	90162268	unknown	GPRIN3	2.656e-13	1.734e-09	0.0006968
rs6532148	90166921	untranslated-3	GPRIN3	6.589e-13	1.74e-09	0.0006968
rs1431548	90168572	untranslated-3	GPRIN3	1.078e-10	0.0008351	0.03055
rs538005	90135739	unknown	GPRIN3	1.983e-10	0.001258	0.1749
rs2736988	90776345	unknown	SNCA	2.094e-10	0.006627	0.05687
rs2737035	90777213	unknown	SNCA	2.094e-10	0.006627	0.05687
rs11733577	89174193	unknown	PPM1K	2.459e-10	3.394e-09	9.054e-13
rs168552	90643144	unknown	SNCA	2.822e-10	0.01156	0.05368
rs894278	90734535	intron	MMRN1	3.026e-10	7.273e-06	NA
rs2583985	90755939	intron	SNCA	3.389e-10	0.006661	0.02271
rs2619361	90757735	intron	SNCA	3.389e-10	0.004987	0.02271
rs2619362	90757845	intron	SNCA	3.389e-10	0.004987	0.02271
rs356174	90630901	unknown	SNCA	4.897e-10	0.02947	0.1582
rs1866995	90731709	intron	MMRN1	6.569e-10	1.031e-06	0.0001333
rs894278	90734535	intron	MMRN1	1.165e-09	5.342e-06	NA
rs2737033	90707947	intron	SNCA	1.277e-09	0.006268	0.05424
rs9307064	90146649	unknown	GPRIN3	1.782e-09	0.01899	0.197
rs9637599	89206230	near-gene-5	PPM1K	2.534e-09	1.825e-09	2.191e-08
rs11931074	90639515	unknown	MMRN1	2.611e-09	0.0002428	0.01135
rs3775423	90657491	intron	MMRN1	2.611e-09	2.52e-05	0.01135
rs3857059	90675238	intron	MMRN1	2.611e-09	2.52e-05	0.01135
rs11944331	90690329	intron	MMRN1	2.611e-09	2.52e-05	0.009669
rs1072587	89168575	unknown	PPM1K	3.154e-09	5.78e-13	1.05e-13
rs6834765	90683990	intron	MMRN1	4.732e-09	2.089e-05	0.003565
rs11931074	90639515	unknown	MMRN1	4.969e-09	0.0001435	0.0006508
rs3775423	90657491	intron	MMRN1	4.969e-09	1.595e-05	0.0006508
rs3857059	90675238	intron	MMRN1	4.969e-09	1.595e-05	0.0006508
rs11944331	90690329	intron	MMRN1	4.969e-09	1.595e-05	0.0005033
rs1866995	90731709	intron	MMRN1	5.864e-09	2.536e-06	5.605e-07
rs958325	89209681	unknown	PPM1K	9.535e-09	4.469e-10	7.158e-09
rs7661312	89201007	intron	PPM1K	1.121e-08	1.816e-09	3.101e-08
rs7656367	89196821	intron	PPM1K	1.172e-08	1.816e-09	3.101e-08
rs168552	90643144	unknown	SNCA	1.219e-08	0.01146	0.1713
rs356174	90630901	unknown	SNCA	1.569e-08	0.03077	0.3699
rs168552	90643144	unknown	SNCA	1.886e-08	0.03586	0.2425
rs7658173	89182518	untranslated-3	PPM1K	2.931e-08	3.837e-09	4.43e-05
rs356174	90630901	unknown	SNCA	3.042e-08	0.04897	0.5177
rs10516849	90789539	unknown	MMRN1	3.159e-08	7.747e-06	2.2e-05
rs6532197	90797301	unknown	MMRN1	3.159e-08	7.747e-06	2.2e-05
rs356220	90641340	unknown	SNCA	3.383e-08	0.1447	0.05151
rs6821589	89192792	intron	PPM1K	3.744e-08	7.721e-08	9.052e-05
rs2869926	89201376	intron	PPM1K	3.744e-08	7.721e-08	9.052e-05
rs881561	89203042	intron	PPM1K	3.744e-08	7.721e-08	9.052e-05
rs6819344	89197519	intron	PPM1K	4.28e-08	7.721e-08	9.052e-05
rs6834765	90683990	intron	MMRN1	4.331e-08	1.477e-05	5.33e-05
rs4423843	89206756	near-gene-5	PPM1K	4.6e-08	8.398e-08	0.0001531

rs10516849	90789539	unknown	MMRN1	6.817e-08	1.662e-05	0.0008057
rs6532197	90797301	unknown	MMRN1	6.817e-08	1.662e-05	0.0008057
rs356220	90641340	unknown	SNCA	1.595e-07	0.1094	0.2813
rs10024717	89219645	unknown	PPM1K	2.146e-07	5.844e-07	9.618e-05
rs34813495	89213077	unknown	PPM1K	2.159e-07	2.308e-06	0.0002057
rs3775465	90827466	intron	MMRN1	2.618e-07	6.313e-07	0.002258
rs7660693	89228383	unknown	PPM1K	2.992e-07	7.022e-06	0.02183
rs28661946	89214757	unknown	PPM1K	3.093e-07	3.35e-09	3.53e-07
rs1442138	90816294	missense	MMRN1	3.912e-07	9.703e-06	2.381e-06
rs10011790	90833421	intron	MMRN1	4.141e-07	0.04635	0.001688
rs518156	90137336	unknown	GPRIN3	5.796e-07	0.08178	0.8506
rs1442138	90816294	missense	MMRN1	5.875e-07	9.961e-06	8.171e-05
rs3775479	90842971	intron	MMRN1	5.935e-07	0.0283	0.002504
rs2737033	90707947	intron	SNCA	7.563e-07	0.04443	0.297
rs3775478	90842840	intron	MMRN1	7.743e-07	7.712e-06	0.002843

Table S2: Significant eQTLs at the 4q31 locus in the Laval dataset and replication in UBC and Groningen.

SNP	SNP Position	SNP Function	Gene Regulated	P Laval	P UBC	P GRNG
rs12498677	146080339	intron	OTUD4	1.058e-08	0.0002261	0.1305
rs7667092	144889757	unknown	BC029578	1.778e-08	7.25E-010	5.95E-011
rs6826862	144899333	unknown	BC029578	3.379e-08	7.25E-010	5.95E-011
rs4287958	144908872	unknown	BC029578	4.828e-08	7.25E-010	5.95E-011
rs1912745	144878737	unknown	BC029578	7.393e-08	2.31E-009	3.48E-011
rs10027246	144856222	unknown	BC029578	1.507e-07	2.50E-008	4.91E-011
rs1545437	144617842	coding-synon	BC029578	1.648e-07	3.38E-008	7.69E-012
rs34714886	145020544	unknown	BC029578	1.85e-07	2.15E-009	3.59E-009
rs1580002	144687660	unknown	BC029578	1.957e-07	1.26E-008	1.24E-011
rs9308176	144993871	unknown	BC029578	1.994e-07	9.24E-010	1.08E-008
rs10030428	144683852	unknown	BC029578	2.036e-07	3.27E-008	1.22E-011
rs35243309	144676269	unknown	BC029578	2.065e-07	9.29E-009	9.71E-012
rs13105102	144740199	unknown	BC029578	2.203e-07	1.26E-008	1.24E-011
rs4835305	144744799	unknown	BC029578	2.203e-07	5.57E-009	9.71E-012
rs6828935	144750266	unknown	BC029578	2.401e-07	3.65E-009	3.50E-012
rs924234	144627952	unknown	BC029578	2.469e-07	5.19E-008	6.54E-012
rs7681914	144629991	unknown	BC029578	2.469e-07	5.19E-008	6.54E-012
rs9685306	144669996	unknown	BC029578	2.57e-07	6.83E-008	5.97E-012
rs1822841	144842638	unknown	BC029578	2.837e-07	2.54E-008	9.62E-011
rs10857415	144829906	unknown	BC029578	2.85e-07	4.49E-008	9.62E-011
rs4240334	144792505	untranslated-3	BC029578	2.973e-07	1.99E-008	2.23E-011
rs9997931	144825754	intron	BC029578	2.973e-07	8.88E-009	7.64E-011
rs7687915	144680954	unknown	BC029578	3.114e-07	8.70E-009	1.75E-011
rs1375989	144648046	unknown	BC029578	3.217e-07	9.50E-010	1.24E-011
rs1450250	144653026	unknown	BC029578	3.217e-07	2.10E-008	1.24E-011
rs13134327	144659795	unknown	BC029578	3.217e-07	9.29E-009	9.71E-012
rs2323190	144665435	unknown	BC029578	3.217e-07	9.29E-009	9.71E-012
rs1375984	144681663	unknown	BC029578	3.217e-07	9.29E-009	9.71E-012
rs1006406	144682863	unknown	BC029578	3.217e-07	9.29E-009	9.71E-012
rs1597445	144689696	unknown	BC029578	3.217e-07	5.57E-009	9.71E-012
rs6537183	144691594	unknown	BC029578	3.217e-07	5.57E-009	9.71E-012
rs12512202	144692401	unknown	BC029578	3.217e-07	5.57E-009	9.71E-012
rs10016223	144698582	unknown	BC029578	3.217e-07	5.57E-009	1.20E-011
rs13129702	144706285	unknown	BC029578	3.217e-07	5.57E-009	9.71E-012
rs4383567	144710026	unknown	BC029578	3.217e-07	6.25E-009	9.71E-012
rs1849136	144833533	unknown	BC029578	5.4e-07	1.99E-008	9.62E-011
rs1489759	145474473	unknown	HHIP	6.687e-07	0.01284	0.2136
rs1828591	145480780	unknown	HHIP	6.687e-07	0.01284	0.2136
rs7655625	145485915	unknown	HHIP	6.687e-07	0.01043	0.2136
rs1375982	144666917	unknown	BC029578	8.758e-07	2.38E-008	1.36E-011
rs6844670	145024799	unknown	BC029578	9.383e-07	1.25E-007	1.70E-006
rs1512282	145431497	unknown	HHIP	1.085e-06	0.03421	0.6693
rs1980057	145485738	unknown	HHIP	1.167e-06	0.01043	0.2136
rs13118928	145486389	unknown	HHIP	1.167e-06	0.01043	0.2121
rs720485	145462588	unknown	HHIP	1.405e-06	0.01043	0.2108
rs6828540	145463231	unknown	HHIP	1.405e-06	0.01043	0.2136
rs11938745	145465768	unknown	HHIP	1.405e-06	0.00962	0.2136
rs1489759	145474473	unknown	HHIP	2.097e-06	0.3051	0.2692
rs1828591	145480780	unknown	HHIP	2.097e-06	0.3051	0.2692

rs7655625	145485915 unknown	HHIP	2.097e-06	0.2619	0.2692
rs9799404	144894260 unknown	BC029578	2.112e-06	2.92E-007	3.76E-006
rs7676614	144616366 intron	BC029578	2.168e-06	5.22E-008	7.46E-012
rs7657795	145009802 unknown	BC029578	2.281e-06	0.0002836	0.01917
rs7679112	145009868 unknown	BC029578	2.641e-06	0.0002836	0.01917
rs1512288	145491281 unknown	HHIP	3e-06	0.009181	0.2272
rs6817273	145492003 unknown	HHIP	3e-06	0.009181	0.2272
rs1980057	145485738 unknown	HHIP	3.207e-06	0.2619	0.2692
rs13118928	145486389 unknown	HHIP	3.207e-06	0.2619	0.2714

Table S3: Significant eQTLs at the 19q13 locus in the Laval dataset and replication in UBC and Groningen.

SNP	SNP Position	SNP Function	Regulated Gene	P Laval	P UBC	P GRNG
rs2191139	42001210	unknown	LOC100505495	9.721e-82	3.22E-067	1.13E-071
rs758643	41991524	unknown	LOC100505495	3.95e-56	2.62E-041	2.48E-049
rs11672691	41985587	unknown	LOC100505495	7.103e-55	9.86E-056	7.24E-054
rs2302188	42085873	missense	CEACAM21	7.741e-51	8.87E-030	3.36E-021
rs887391	41985624	unknown	LOC100505495	2.09e-49	4.20E-040	2.01E-046
rs714106	42083849	missense	CEACAM21	1.074e-45	3.15E-027	1.23E-019
rs4803481	42066556	unknown	CEACAM21	1.231e-45	1.22E-027	3.60E-019
rs28605876	41931454	untranslated-3	B3GNT8	1.718e-40	1.97E-012	5.12E-014
rs284662	41932275	missense	B3GNT8	4.694e-39	2.97E-011	2.81E-014
rs10409485	41920765	intron	B3GNT8	2e-38	1.51E-011	5.69E-015
rs1473248	41923314	intron	B3GNT8	2e-38	1.51E-011	5.69E-015
rs284652	41928652	coding-synon	B3GNT8	2e-38	1.51E-011	5.69E-015
rs4674	41930396	coding-synon	B3GNT8	2e-38	2.88E-011	7.82E-014
rs284663	41932612	coding-synon	B3GNT8	2e-38	3.06E-011	4.93E-014
rs11670757	41913431	intron	B3GNT8	2.775e-38	2.73E-011	4.12E-014
rs4803465	41918158	intron	B3GNT8	2.775e-38	1.51E-011	5.69E-015
rs11880539	41918949	intron	B3GNT8	2.775e-38	1.51E-011	8.10E-014
rs284661	41932120	coding-synon	B3GNT8	2.775e-38	3.06E-011	6.30E-013
rs12973063	42009673	unknown	LOC100505495	1.079e-37	5.49E-036	2.73E-051
rs4803480	42066279	unknown	CEACAM21	1.267e-37	4.45E-033	4.60E-018
rs11882144	42072602	unknown	CEACAM21	1.267e-37	7.01E-035	8.25E-019
rs2231940	41944237	missense	B3GNT8	1.605e-37	2.21E-011	1.70E-012
rs9304593	41942875	intron	B3GNT8	2.211e-37	2.21E-011	3.08E-012
rs13344547	41945848	near-gene-5	B3GNT8	2.211e-37	2.37E-011	1.70E-012
rs3745290	41890003	near-gene-3	B3GNT8	4.201e-37	4.73E-012	2.64E-014
rs878082	41897674	intron	B3GNT8	4.201e-37	8.01E-012	2.64E-014
rs10853751	41903220	missense	B3GNT8	4.354e-37	9.72E-012	1.78E-015
rs892044	41903671	near-gene-5	B3GNT8	4.354e-37	9.72E-012	1.78E-015
rs3810174	41904165	intron	B3GNT8	4.354e-37	9.72E-012	1.78E-015
rs12602	41889748	coding-synon	B3GNT8	1.341e-36	3.60E-011	2.67E-015
rs4141218	42081907	near-gene-5	CEACAM21	5.406e-34	3.04E-031	2.28E-017
rs17318596	41937095	near-gene-3	B3GNT8	3.102e-33	2.52E-009	1.46E-013
rs4802122	42047924	unknown	CEACAM21	2.588e-28	5.99E-022	8.86E-011
rs7255742	42049821	unknown	CEACAM21	4.744e-28	1.34E-024	2.70E-012
rs10412700	42020942	unknown	LOC100505495	2.305e-26	4.70E-034	1.59E-038
rs12978323	41756038	intron	AXL	3.761e-26	4.90E-011	5.78E-008
rs7250883	41760033	intron	AXL	6.755e-26	8.06E-009	5.33E-008
rs12980267	41765229	intron	AXL	6.755e-26	8.06E-009	5.33E-008
rs1946612	41767987	near-gene-3	AXL	6.755e-26	8.06E-009	5.33E-008
rs3760659	41769104	intron	AXL	6.755e-26	8.06E-009	6.77E-008
rs371500	40480923	intron	ZNF780A	5.235e-25	3.77E-014	3.56E-010
rs337799	40591182	intron	ZNF780A	2.072e-24	1.53E-021	2.83E-016
rs10402038	40637615	unknown	ZNF780A	2.072e-24	3.03E-020	2.83E-016
rs16974093	40650002	unknown	ZNF780A	2.072e-24	3.03E-020	1.65E-016
rs11668699	42031592	unknown	LOC100505495	3.751e-24	1.65E-030	5.14E-031
rs3826716	42035716	unknown	LOC100505495	3.751e-24	4.35E-032	2.16E-030
rs11668699	42031592	unknown	CEACAM21	6.066e-24	1.20E-017	2.05E-009
rs3826716	42035716	unknown	CEACAM21	6.066e-24	2.58E-018	3.66E-009
rs3816052	41773797	intron	AXL	9.161e-24	1.32E-013	2.61E-009
rs2304230	41774064	intron	AXL	9.161e-24	1.32E-013	6.32E-010

rs3745295	41785282	intron	AXL	9.161e-24	1.32E-013	6.32E-010
rs3752705	41798111	intron	AXL	9.161e-24	1.32E-013	6.32E-010
rs12980480	41804655	intron	AXL	9.161e-24	1.32E-013	6.32E-010
rs7248164	41807700	intron	AXL	9.161e-24	1.32E-013	2.61E-009
rs2304232	41762525	intron	AXL	1.301e-23	1.10E-011	2.63E-008
rs2304231	41762670	intron	AXL	1.301e-23	1.10E-011	2.03E-008
rs4591267	41765407	intron	AXL	1.9e-23	1.15E-010	5.71E-009
rs11879090	41917260	intron	B3GNT8	3.103e-23	1.31E-009	9.90E-012
rs284653	41928701	intron	B3GNT8	3.103e-23	1.41E-009	9.90E-012
rs11879435	41757603	intron	AXL	4.748e-23	7.53E-012	2.67E-007
rs459091	40561175	intron	ZNF780A	5.342e-23	5.51E-014	1.44E-010
rs234346	40467676	unknown	ZNF780A	5.784e-23	5.51E-014	2.67E-010
rs234370	40477465	intron	ZNF780A	5.784e-23	5.51E-014	1.44E-010
rs8105066	40489470	unknown	ZNF780A	5.784e-23	5.51E-014	1.44E-010
rs426950	40562386	near-gene-5	ZNF780A	5.784e-23	5.73E-015	1.44E-010
rs421533	40579468	intron	ZNF780A	5.784e-23	3.64E-014	1.44E-010
rs11882467	41754217	intron	AXL	7.435e-23	1.56E-008	2.00E-007
rs170901	40532800	unknown	ZNF780A	7.923e-23	1.56E-015	8.26E-011
rs12974420	41798660	intron	AXL	8.448e-23	6.97E-013	6.32E-010
rs4577216	40569551	unknown	ZNF780A	9.677e-23	4.81E-016	1.29E-010
rs24136	40596207	intron	ZNF780A	1.05e-22	4.06E-014	1.29E-010
rs337789	40599603	unknown	ZNF780A	1.05e-22	4.06E-014	1.29E-010
rs1327189	40627509	unknown	ZNF780A	4.12e-22	6.72E-012	2.80E-009
rs1998931	40628064	unknown	ZNF780A	4.12e-22	5.33E-011	2.80E-009
rs17761603	40641794	unknown	ZNF780A	4.12e-22	6.72E-012	2.80E-009
rs2767607	40611153	unknown	ZNF780A	4.747e-22	6.21E-012	2.80E-009
rs11542840	40480501	unknown	ZNF780A	4.783e-22	1.82E-013	2.80E-009
rs234324	40502919	near-gene-5	ZNF780A	4.952e-22	2.45E-013	1.97E-009
rs183504	40508115	intron	ZNF780A	6.459e-22	2.45E-013	2.80E-009
rs2767604	40609975	unknown	ZNF780A	6.459e-22	6.72E-012	2.80E-009
rs2191241	42103552	unknown	LOC100505495	2.745e-21	1.31E-012	9.88E-019
rs234325	40502943	untranslated-5	ZNF780A	7.254e-21	2.62E-013	1.97E-009
rs10412700	42020942	unknown	CEACAM21	1.285e-20	4.90E-017	1.29E-008
rs1056854	41809995	coding-synon	AXL	1.36e-20	2.24E-012	1.11E-009
rs1076362	42091163	intron	CEACAM21	5.545e-20	3.81E-009	4.11E-007
rs1865053	41814412	near-gene-5	AXL	6.876e-20	3.62E-011	1.21E-008
rs7245398	41810644	intron	AXL	9.23e-20	3.62E-011	5.00E-008
rs234309	40517991	intron	ZNF780A	1.256e-19	1.85E-012	6.71E-009
rs234353	40529219	unknown	ZNF780A	1.369e-19	1.85E-012	6.71E-009
rs2191241	42103552	unknown	CEACAM21	1.474e-19	1.53E-016	2.79E-010
rs4802122	42047924	unknown	LOC100505495	2.764e-19	3.00E-024	2.65E-023
rs12984160	41815713	near-gene-5	AXL	6.605e-19	1.64E-010	1.21E-008
rs2005832	41983822	unknown	LOC100505495	9.981e-19	1.76E-032	1.17E-018
rs12981943	42019187	unknown	LOC100505495	1.436e-18	1.52E-018	7.32E-018
rs12610088	42020112	unknown	LOC100505495	4.924e-18	2.60E-018	5.46E-018
rs2241719	41829581	3' UTR	AXL	6.422e-18	2.19E-009	2.91E-008
rs2241718	41829606	untranslated-3	AXL	8.234e-18	1.47E-009	1.30E-008
rs714106	42083849	missense	LOC100505495	1.045e-17	5.62E-020	8.32E-018
rs4803481	42066556	unknown	LOC100505495	1.321e-17	4.64E-020	8.74E-019
rs12985162	41831507	unknown	AXL	1.428e-17	4.33E-009	2.85E-008
rs1865051	41824885	intron	AXL	2.371e-17	1.92E-010	1.30E-008
rs7257310	41827120	intron	AXL	2.371e-17	1.92E-010	1.21E-008
rs6957	41830606	untranslated-3	AXL	2.371e-17	1.92E-010	4.65E-008
rs758643	41991524	unknown	CEACAM21	2.804e-17	5.81E-011	3.15E-005
rs1029804	42107207	unknown	LOC100505495	6.567e-17	1.31E-015	2.07E-019
rs8109167	41823167	intron	AXL	7.372e-17	1.92E-010	1.58E-008

rs1046909	41882712	intron	B3GNT8	7.383e-17	4.94E-006	7.59E-005
rs12973063	42009673	unknown	CEACAM21	7.93e-17	8.59E-016	2.49E-006
rs12983047	41834499	unknown	AXL	1.145e-16	2.57E-009	2.30E-007
rs8105161	41839631	intron	AXL	1.145e-16	3.56E-009	7.08E-008
rs4803464	41918054	intron	B3GNT8	6.912e-16	7.84E-007	1.30E-006
rs2302188	42085873	missense	LOC100505495	1.12e-15	2.22E-018	8.11E-015
rs2191139	42001210	unknown	CEACAM21	1.927e-15	1.96E-008	8.38E-005
rs7255742	42049821	unknown	LOC100505495	1.967e-15	2.78E-013	2.14E-019
rs1029804	42107207	unknown	CEACAM21	5.78e-15	6.00E-011	1.99E-006
rs2304234	41748753	intron	AXL	2.155e-14	2.28E-005	0.000261
rs2241713	41869468	intron	B3GNT8	1.314e-13	4.50E-006	4.56E-005
rs887391	41985624	unknown	CEACAM21	3.053e-13	1.61E-007	0.003628
rs4803457	41861359	intron	B3GNT8	3.437e-13	4.66E-006	6.21E-005
rs11083616	41865643	intron	B3GNT8	3.437e-13	4.66E-006	2.06E-005
rs11666933	41862253	intron	B3GNT8	3.809e-13	4.30E-006	7.53E-005
rs4803449	41734666	intron	AXL	2.241e-12	0.001843	0.0007176
rs4090570	42116126	unknown	CEACAM4	4.026e-12	0.001642	0.006108
rs2191240	42103501	unknown	LOC100505495	4.072e-12	1.52E-011	6.78E-019
rs2317314	42062307	unknown	CEACAM21	4.821e-12	1.07E-012	5.91E-006
rs2191240	42103501	unknown	CEACAM21	5.156e-12	4.59E-016	1.04E-007
rs12610088	42020112	unknown	CEACAM21	5.506e-12	2.45E-009	0.0003046
rs12981943	42019187	unknown	CEACAM21	6.146e-12	1.59E-009	0.0003289
rs11672691	41985587	unknown	CEACAM21	1.644e-11	1.03E-007	0.003572
rs16975338	42058025	unknown	CEACAM21	1.911e-11	5.66E-014	2.77E-008
rs474481	40665924	unknown	ZNF780A	3.629e-11	2.08E-012	1.11E-005
rs34466648	40433042	coding-synon	ZNF780A	4.876e-11	0.01291	0.06617
rs2231738	40450048	unknown	ZNF780A	4.932e-11	9.89E-005	0.001084
rs3760920	40442558	unknown	ZNF780A	5.272e-11	0.01144	0.01882
rs3745938	42071167	unknown	LOC100505495	5.314e-11	1.31E-006	1.26E-006
rs7247842	42083673	coding-synon	LOC100505495	5.314e-11	1.35E-005	1.26E-006
rs740591	42084586	intron	LOC100505495	5.314e-11	1.35E-005	1.26E-006
rs8106269	40447257	unknown	ZNF780A	6.262e-11	0.0002237	0.001845
rs8109718	40447840	unknown	ZNF780A	6.262e-11	0.0002237	0.002637
rs12462093	41984128	unknown	LOC100505495	6.346e-11	1.85E-019	5.37E-014
rs6508987	42120778	unknown	CEACAM4	7.485e-11	0.0003868	0.008244
rs7260507	41947625	near-gene-5	B3GNT8	1.389e-10	5.54E-006	0.0002377
rs7257657	42082326	near-gene-5	LOC100505495	2.519e-10	8.16E-005	0.001288
rs10421785	42068030	unknown	LOC100505495	2.901e-10	2.69E-005	0.001288
rs10421584	42067978	unknown	LOC100505495	2.949e-10	2.69E-005	0.001288
rs758350	42085162	intron	LOC100505495	2.949e-10	2.69E-005	0.0007341
rs2215324	42120283	unknown	CEACAM4	2.971e-10	0.001225	0.009485
rs8109627	41822986	intron	AXL	3.264e-10	2.51E-005	0.0002727
rs3760919	40442993	unknown	ZNF780A	4.313e-10	0.006192	0.002752
rs2241715	41856886	intron	B3GNT8	5.767e-10	0.0001956	0.007973
rs2241714	41869392	missense	B3GNT8	6.596e-10	0.0001942	0.007339
rs1800469	41860296	near-gene-3	B3GNT8	6.75e-10	0.0001715	0.009191
rs4802113	41740895	intron	AXL	7.143e-10	0.004508	0.01743
rs7246525	41737996	intron	AXL	2.079e-09	0.0111	0.009842
rs4803455	41851509	intron	B3GNT8	2.146e-09	0.0004879	0.005567
rs4802111	41734059	intron	AXL	3.275e-09	0.004942	0.022
rs4803448	41734560	intron	AXL	5.425e-09	0.006571	0.01937
rs16975338	42058025	unknown	LOC100505495	6.263e-09	3.97E-010	1.10E-015
rs10401344	41192417	intron	NUMBL	9.554e-09	3.90E-007	3.16E-008
rs2355720	40440064	intron	ZNF780A	1.601e-08	0.02128	0.1223
rs3786938	40441622	near-gene-5	ZNF780A	1.607e-08	0.04681	0.1051
rs7249890	40441024	near-gene-5	ZNF780A	1.619e-08	0.04495	0.07561

rs2215324	42120283	unknown	CEACAM21	2.031e-08	6.66E-005	0.01062
rs186816	40594660	intron	ZNF780A	4.436e-08	1.30E-007	1.85E-007
rs1041992	42132273	coding-synon	LOC100505495	5.173e-08	NA	0.6647
rs6508987	42120778	unknown	CEACAM21	5.754e-08	2.41E-005	0.04661
rs371500	40480923	intron	ZNF780A	6.809e-08	4.93E-007	0.0113
rs337799	40591182	intron	ZNF780A	7.343e-08	1.19E-008	0.0001668
rs10402038	40637615	unknown	ZNF780A	7.343e-08	7.45E-009	0.0001668
rs16974093	40650002	unknown	ZNF780A	7.343e-08	7.45E-009	0.0002193
rs12461895	41848347	intron	B3GNT8	1.543e-07	0.006515	0.128
rs4141218	42081907	near-gene-5	LOC100505495	1.633e-07	4.43E-011	5.46E-011
rs8106269	40447257	unknown	ZNF780A	2.264e-07	8.00E-006	0.03561
rs8109718	40447840	unknown	ZNF780A	2.264e-07	8.00E-006	0.03206
rs3901991	42128922	intron	LOC100505495	2.292e-07	0.2419	0.6952
rs4803471	41963749	unknown	LOC100505495	2.295e-07	3.32E-010	0.001875
rs3745205	40948202	intron	SERTAD3	3.362e-07	8.71E-006	0.0003802
rs929502	42130776	intron	CEACAM4	3.37e-07	0.0009423	0.07069
rs8111595	41989138	unknown	LOC100505495	3.809e-07	0.01018	NA
rs890934	41227968	intron	NUMBL	4.214e-07	4.97E-005	6.45E-009
rs234346	40467676	unknown	ZNF780A	4.227e-07	1.25E-007	0.002509
rs234370	40477465	intron	ZNF780A	4.227e-07	1.25E-007	0.003179
rs8105066	40489470	unknown	ZNF780A	4.227e-07	1.25E-007	0.003179
rs426950	40562386	near-gene-5	ZNF780A	4.227e-07	3.12E-008	0.003179
rs421533	40579468	intron	ZNF780A	4.227e-07	2.07E-008	0.003179
rs459091	40561175	intron	ZNF780A	4.361e-07	1.25E-007	0.003179
rs2887292	40947448	coding-synon	SERTAD3	5.667e-07	8.71E-006	0.0003802
rs35314597	40947751	coding-synon	SERTAD3	5.667e-07	8.71E-006	0.0003802
rs10421785	42068030	unknown	CEACAM21	6.709e-07	0.001786	0.0002254
rs2316973	41953928	unknown	LOC100505495	6.818e-07	1.33E-005	0.0004636
rs186816	40594660	intron	ZNF780A	6.953e-07	2.09E-007	0.006835

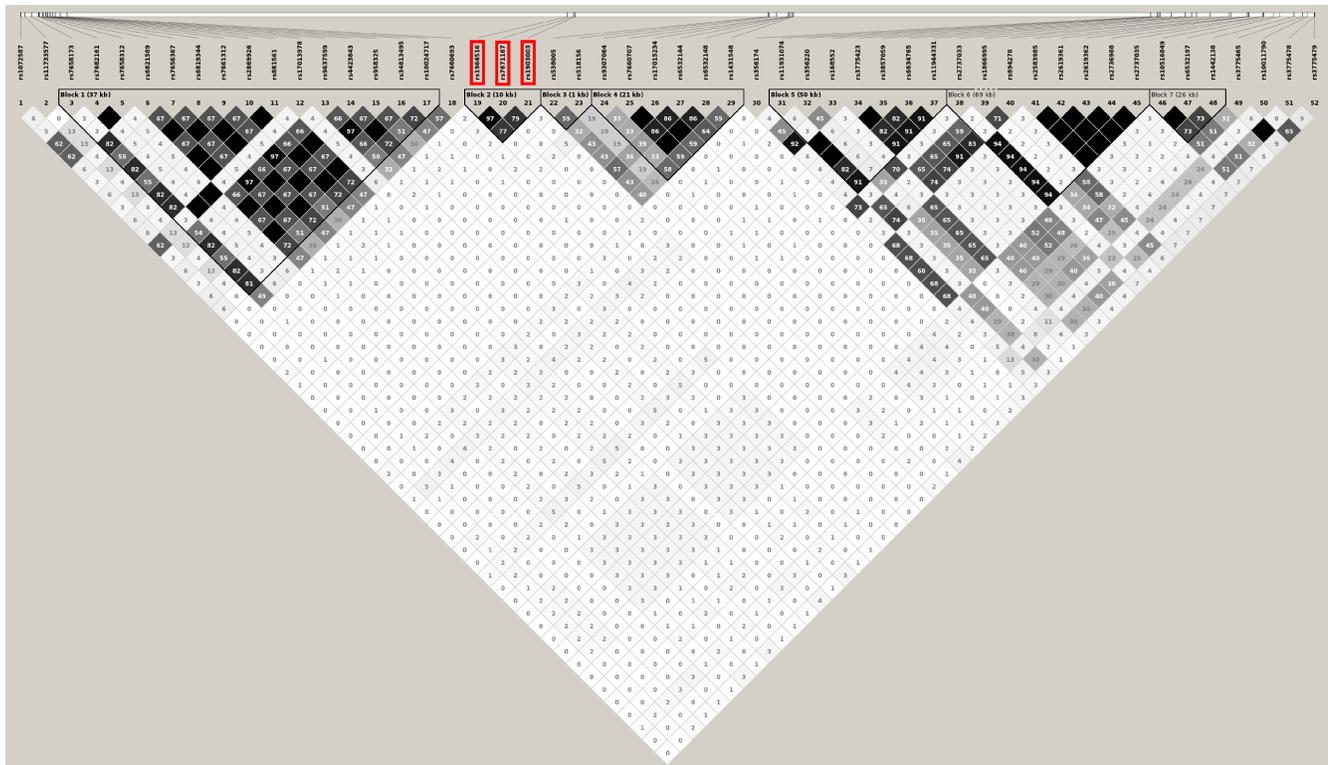


Figure S1: Linkage disequilibrium plot of significant SNPs on the 4q22 locus in the 1000 Genome Project. The white horizontal bar on the upper part of the figure illustrates the location of SNPs on a physical scale. LD values (r^2) are indicated in each box. The color of the squares illustrate the strength of pairwise r^2 values on a black and white scale where black indicates perfect LD ($r^2 = 1$) and white indicates perfect equilibrium ($r^2 = 0$). Red rectangles are SNPs previously associated with COPD (Table 2). The genotypes are from the 1000 Genome Project interim phase1 release (2010/11/23).

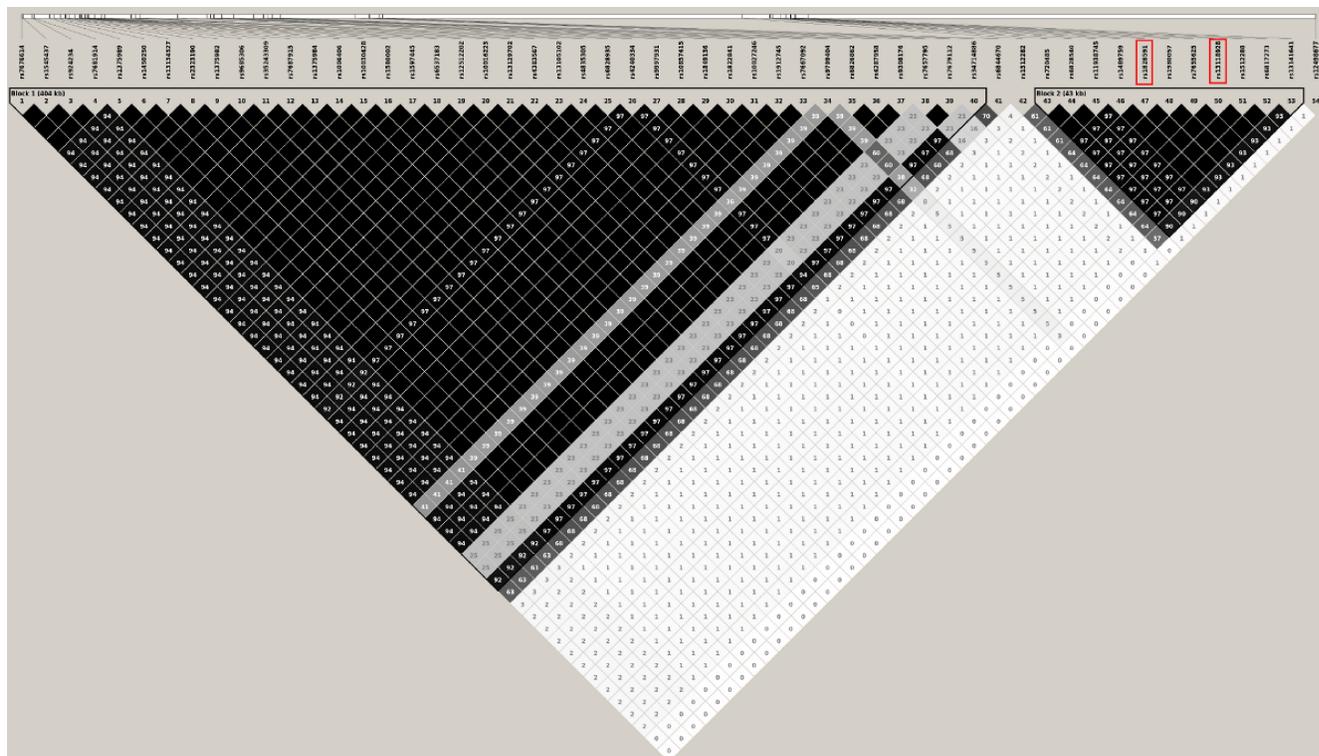


Figure S2: Linkage disequilibrium plot of significant SNPs on the 4q31 locus in the 1000 Genome Project. The white horizontal bar on the upper part of the figure illustrates the location of SNPs on a physical scale. LD values (r^2) are indicated in each box. The color of the squares illustrate the strength of pairwise r^2 values on a black and white scale where black indicates perfect LD ($r^2 = 1$) and white indicates perfect equilibrium ($r^2 = 0$). Red rectangles are SNPs previously associated with COPD (Table 2). The genotypes are from the 1000 Genome Project interim phase1 release (2010/11/23).

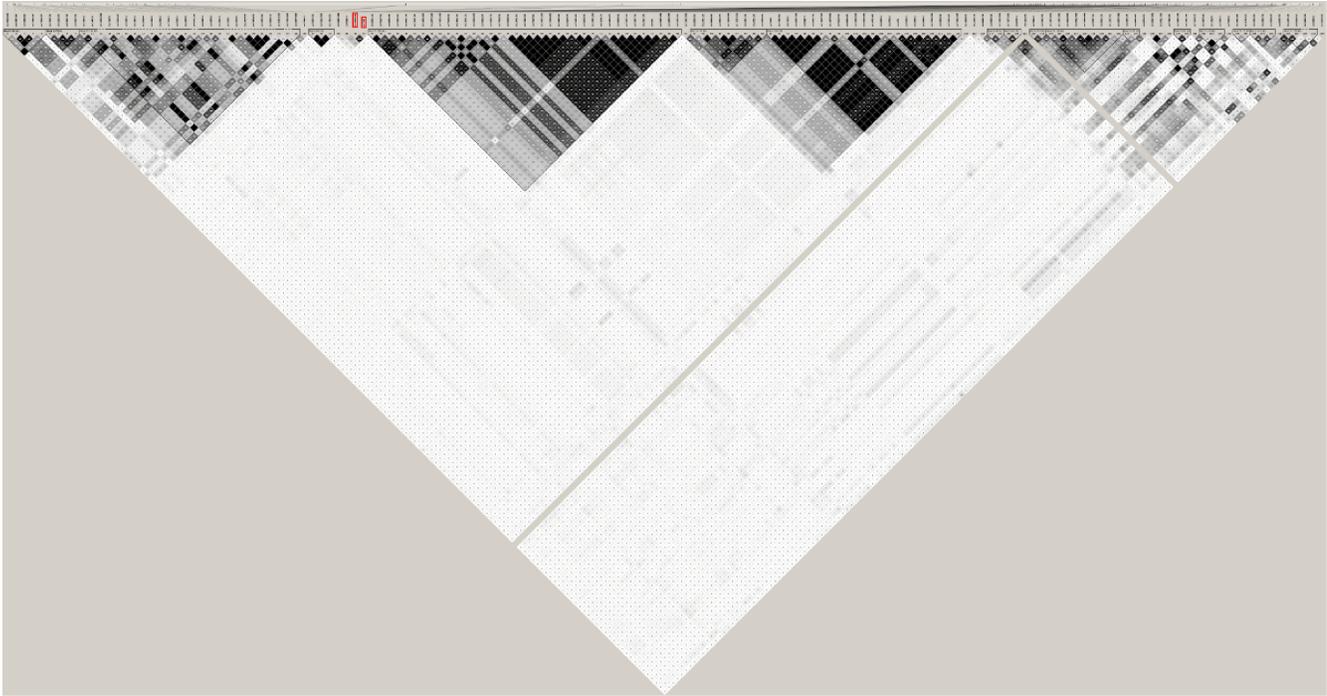


Figure S3: Linkage disequilibrium plot of significant SNPs on the 19q13 locus in the 1000 Genome Project. The white horizontal bar on the upper part of the figure illustrates the location of SNPs on a physical scale. LD values (r^2) are indicated in each box. The color of the squares illustrate the strength of pairwise r^2 values on a black and white scale where black indicates perfect LD ($r^2 = 1$) and white indicates perfect equilibrium ($r^2 = 0$). Red rectangles are SNPs previously associated with COPD (Table 2). The genotypes are from the 1000 Genome Project interim phase1 release (2010/11/23).

Chapitre 3 : Conclusion

La maladie pulmonaire obstructive chronique (MPOC) est une maladie qui affecte les voies respiratoires et les poumons. Elle se caractérise par une réduction lente et progressive des fonctions respiratoires⁹. Le principal facteur de risque de la maladie est le tabagisme, mais des facteurs environnementaux, une déficience en alpha-1 antitrypsine et des variations génétiques peuvent influencer le développement de la maladie^{1,11,22}. Un exemple frappant de l'importance de la génétique dans cette maladie est le fait que seulement 15 à 20 % des fumeurs développent la maladie, même si les fumeurs représentent plus de 85 % des personnes atteintes de MPOC¹. Récemment, quatre régions chromosomiques furent associées à la MPOC lors d'études génomique de grande envergure : 4q22, 4q31, 15q25, 19q13^{59,63,64}. Pour chacune de ces régions, certains gènes sont soupçonnés comme pouvant avoir un impact sur le développement de la MPOC. Par contre, aucune étude n'a jusqu'à maintenant vérifié si ces gènes sont bien les gènes causaux dans ces régions. Dans le but d'identifier les gènes causaux dans trois de ces quatre locus, nous avons élaboré une étude d'eQTLs dans les poumons. Celle-ci consiste à identifier les polymorphismes influençant l'expression des gènes dans les poumons à l'aide des génotypes et des valeurs d'expression d'un grand nombre de sujets. Dans ce mémoire, les trois régions analysées furent 4q22, 4q31 et 19q13. Le locus 15q25.1 fut étudié dans le cadre d'une autre étude au laboratoire. L'influence des polymorphismes sur l'expression des gènes fut vérifiée dans chaque région. Cette approche permet de déterminer si un SNP a un impact sur l'expression d'un gène situé dans son voisinage. Les populations à l'étude furent trois cohortes de patients opérés pour diverses conditions pulmonaires dont nous avons un échantillon sanguin et un échantillon pulmonaire non tumoral.

Le chapitre 2 de ce mémoire présente les eQTLs identifiés dans les trois régions chromosomiques d'intérêts. Les eQTLs pulmonaires furent identifiés dans une première cohorte et ensuite répliqués dans deux autres jeux de données. Notre première cohorte provenait de l'IUCPQ. Les deux autres provenaient de l'Université de la Colombie-Britannique et de l'Université de Groningen. Toutes les cohortes étaient composées de patients opérés aux poumons. Cette étude nous a

permis d'identifier un grand nombre d'eQTLs dans les trois régions visées. De plus, ces résultats nous ont permis d'identifier de nouveaux gènes ayant possiblement un impact dans le développement de la MPOC. Les rôles d'une partie de ces gènes sont encore peu connus, ce qui rend l'interprétation biologique des résultats difficiles. Par contre, cette étude seule ne peut nous permettre d'affirmer que ces gènes sont directement impliqués dans le développement de la MPOC. De nouvelles études sont nécessaires pour déterminer dans un premier temps le rôle de ces gènes dans les poumons et vérifier dans un deuxième temps si ces gènes sont associés au développement de la MPOC. Les résultats pour les régions 4q22 et 19q13 ne nous permettent pas d'affirmer que les gènes candidats sont bel et bien les gènes causals, mais les résultats tendent vers cette conclusion. Seul des associations dans le gène *HHIP* furent significatives, mais celles-ci n'étaient pas présentes dans la cohorte de Groningen. Nos résultats ne peuvent être comparés à aucune autre étude, puisque c'est la première étude d'eQTLs pulmonaire à se concentrer sur ces régions d'intérêts.

Les associations significatives détectées dans le locus 4q22 régulaient les gènes *GPRIN3*, *MMRN1*, *PPM1K* et *SNCA*. Les SNPs précédemment associés à la MPOC dans ce locus n'étaient disponibles dans nos données et aucun autre SNP ne fut associé à l'expression du gène *FAM13A*. Cependant, 5 variations présentent dans nos données étaient en LD avec ceux de la littérature ($r^2 > 0,5$). La meilleure association fut observée entre trois SNPs en LD complet (rs2045517, rs2869967, rs2869966) et le gène *FAM13A* ($p = 4.1 \times 10^{-5}$). Ces résultats nous permettent de supporter mais non de confirmer que le gène *FAM13A* est le gène causal dans cette région.

Pour le locus 4q31, quelques SNPs précédemment associés à la MPOC (rs1828591^{59,64}, rs13118928^{59,64}) furent associés dans notre étude avec l'expression du gène *HHIP* dans les cohortes de Laval et UBC, mais pas dans la cohorte de Groningen. Des associations furent également détectées avec les gènes *OTUD4* et *BC029578*, mais seul ceux régulant *BC029578* répliquèrent dans UBC et Groningen. Ces résultats semblent démontrer que *HHIP* pourrait être le gène

causal de cette région, mais la variance expliquée par les variations génétiques semble faible.

En ce qui concerne le locus 19q13, des associations SNP-gène ont été identifiées avec huit gènes et ce dans les trois cohortes : *AXL*, *B3GNT8*, *CEACAM21*, *CEACAM4*, *LOC100505495*, *NUMBL*, *SERTAD3*, *ZNF780A*. Un des deux SNPs précédemment associé à la MPOC (rs7937⁶³) était présent dans nos données. Celui-ci est en LD avec le deuxième SNP, rs2604894⁶³ ($r^2 = 0.82$). Aucun test statistique ne fut significatif avec le SNP rs7937. La meilleure association observée avec cette variation fut avec le gène *NUMBL* ($p = 0.0187$). La meilleure association avec l'un des gènes suspectés était rs4803369-*EGLN2* ($p = 8.9 \times 10^{-7}$). Cette variation était en faible LD avec le SNP rs7937 ($r^2 = 0.33$). *EGLN2* pourrait donc être le gène en cause dans cette région.

Cette étude comporte certaines limitations. Tout d'abord, il existe des différences cliniques entre les cohortes. La cohorte de Groningen est composée de sujets plus jeunes avec des fonctions pulmonaires plus altérées. L'indice de masse corporelle de ces sujets était également plus faible. Pourtant, cette cohorte possède le plus haut taux de non-fumeur et la plus faible consommation tabagique. Puisque la MPOC est une maladie complexe, il est important d'analyser des sujets avec des phénotypes précis⁷⁴. Les résultats seront différents entre une cohorte de personnes souffrant d'emphysème versus une autre souffrant de bronchite chronique. Les différences phénotypiques entre nos trois cohortes peuvent expliquer une partie des associations qui ne répliquent pas. Le fait que plusieurs SNPs de la littérature n'étaient pas disponibles dans nos données limite notre pouvoir de confirmer ou d'infirmer leur rôle dans l'expression des gènes. La présence de SNPs en fort LD aurait pu nous permettre de contourner ce problème, mais cela ne fut pas le cas pour la région 19q13.

Nos résultats sont très complémentaires à ceux des criblages génomiques par associations. Ces dernières études ont identifiés des variations génétiques associées à la maladie. De notre côté, nous avons identifié des variations génétiques dans les mêmes régions associés à l'expression des gènes dans les

poumons. Il est tout à fait possible que les variations génétiques associées à la maladie n'aient aucun impact sur l'expression des gènes dans les poumons. Ils peuvent être associés à l'expression de gènes présents dans d'autres tissus important pour la MPOC, puisque les eQTLs peuvent être spécifique à un tissu⁷⁵. Il se pourrait également que les variations précédemment associées soient des *trans*-eQTLs. Cette option ne fut pas abordée dans le cadre cette étude. Une étude d'eQTLs sur le crachat humain a fait des observations similaires pour ce qui est des gènes causaux hypothétiques dans la MPOC⁷⁶. Cette étude n'avait détecté aucun eQTL à plus ou moins 200 kb des gènes *FAM13A* et *HHIP* qui sont les gènes soupçonnés dans les régions 4q22 et 4q31. Ils n'ont cependant pas analysés le locus 19q13 qui fut associé à la MPOC en 2012.

Les résultats de ce mémoire apportent de nouvelles cibles pour les futures études génomiques sur la MPOC. Nous avons démontré que plusieurs gènes sont régulés par des variations génétiques dans les poumons. Très peu d'informations sont disponibles sur le fonctionnement de certains de ces gènes. Par exemple, des associations furent détectées dans le gène *GPRIN3* qui se trouve à 187 kb du gène *FAM13A*, mais peu d'information est disponible sur le rôle biologique de *GPRIN3*. Tous les gènes dans cette situation sont maintenant de nouvelles cibles pour des études futures (*OTUD4*, *BC029578*, *B3GNT8*, *CEACAM21*, *CEACAM4*, *LOC100505495*, *NUMBL*, *ZNF780A*). Il est essentiel de connaître le fonctionnement de ces gènes, leur rôle dans les poumons et de déterminer s'il existe un lien entre ces gènes et la MPOC. Des analyses bio-informatiques des domaines formant les protéines codées par ces gènes seraient appropriées pour avoir une idée de la famille de gènes auxquels ils appartiennent. Des études en laboratoire avec des souris transgéniques pourraient également être effectuées, mais cela nécessite des investissements importants en temps. Des études cas-témoins avec les gènes nouvellement identifiés sont nécessaires pour vérifier s'ils ont un impact sur le développement de la MPOC. Ces études pourraient être réalisées avec les techniques de génotypages classiques, ou avec le séquençage de nouvelles générations. Il faudrait vérifier si les génotypes obtenus sont associés au risque de développer la MPOC, ou avec un autre phénotype d'intérêt comme la

surproduction de mucus.

Nos résultats n'ont aucune implication clinique pour l'instant, mais ils nous ont permis d'identifier de nouvelles cibles potentiellement thérapeutiques. Avant de pouvoir transposer ces nouvelles cibles vers le traitement ou la prévention de la MPOC, des études supplémentaires doivent être effectuées pour vérifier s'il existe un lien avec la MPOC, pour comprendre leur fonctionnement et ultimement pour développer des médicaments. Il s'écoulera encore plusieurs années avant que la clinique profite de cette recherche.

Il devient également de plus en plus important de savoir dans quel tissu un gène est exprimé. La littérature scientifique déborde de gènes associés à la MPOC^{23,24}, mais la manière dont ceux-ci affectent le développement de cette maladie reste encore peu connue. Puisque les eQTLs peuvent être spécifique à un tissu⁷⁵, il sera particulièrement important de connaître les tissus exprimant les gènes identifiés dans ces études afin de développer de nouveaux médicaments. De nouvelles études d'eQTLs seront nécessaires dans d'autres tissus importants pour cette maladie afin de mieux comprendre le rôle des variations génétiques précédemment détectés.

Il est nécessaire d'effectuer de nouveaux criblages génomiques par association avec un grand nombre d'individus et des phénotypes précis pour identifier de nouveaux loci de risque. Ce genre d'étude nécessite le rassemblement de plusieurs cohortes afin d'élargir le nombre de sujet total. Cette étape est déjà en cours. L'*International COPD Genetics Consortium* fut formé spécifiquement dans ce but⁷⁴. Ce consortium regroupe plusieurs cohortes axées sur la génétique de la MPOC. Des critères phénotypiques précis furent utilisés pour la sélection des sujets. Les résultats statistiques de chaque cohorte seront par la suite combinés dans une méta-analyse dans le but d'identifier de nouveaux locus associés à la MPOC.

Les nouvelles technologies de séquençage à haut débit doivent également être mises à contribution dans la recherche génétique sur la MPOC. Celles-ci

permettraient de séquencer des gènes précédemment associés à la MPOC pour identifier toutes les variations génétiques présentes. Ces données pourraient par la suite être mises en commun avec des échantillons de tissus liés à la maladie pour identifier les variations causales dans les gènes. Contrairement aux études effectuées avec des puces à ADN comme la nôtre, toutes les variations génétiques présentes dans le gène visé seraient alors testées.

Malgré que les études d'eQTLs nous permettent de mieux comprendre l'impact des variations génétiques sur l'expression ces gènes, ces nouvelles informations sont encore trop limitées pour nous permettre de les appliquer à la MPOC. De plus, les tissus nécessaires pour réaliser ce type d'étude sont extrêmement difficiles à se procurer en quantité suffisante. Dans le but d'optimiser les découvertes réalisées lors de cette étude, des études bio-informatique et fonctionnelles seront nécessaires pour bien comprendre le fonctionnement de chaque gène identifié et pour vérifier si ces gènes peuvent avoir un impact sur la MPOC. Pour savoir si ces gènes peuvent être reliés à la MPOC, des études cas-témoins avec des gènes candidats seraient également nécessaires. Toutes ces nouvelles analyses pourraient ouvrir la voie à une meilleure compréhension des gènes associés à la MPOC.

Bibliographie

1. Rabe, K. F. *et al.* Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *American Journal of Respiratory and Critical Care Medicine* **176**, 532–555 (2007).
2. Patel, B. D. *et al.* Airway wall thickening and emphysema show independent familial aggregation in chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine* **178**, 500–505 (2008).
3. Kohansal, R. *et al.* The natural history of chronic airflow obstruction revisited: an analysis of the Framingham offspring cohort. *American Journal of Respiratory and Critical Care Medicine* **180**, 3–10 (2009).
4. Halbert, R. J. *et al.* Global burden of COPD: systematic review and meta-analysis. *The European Respiratory Journal* **28**, 523–532 (2006).
5. Who, W. H. O. *World Health Statistics 2008*. 212 (2008).
6. Statistics Canada Canadian Community Health Survey. *Active Living and the Working Population* **1**, 1–8 (2003).
7. Canada, P. H. A. O. Life and Breath: Respiratory Disease in Canada. *Transplantation* **115** (2007).
8. Mittmann, N. *et al.* The cost of moderate and severe COPD exacerbations to the Canadian healthcare system. *Respiratory Medicine* **102**, 413–421 (2008).
9. GICOLD. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease revised 2011. (2011).
10. Hogg, J. C. Pathophysiology of airflow limitation in chronic obstructive pulmonary disease. *Lancet* **364**, 709–721 (2004).
11. Ganrot, P. O., Laurell, C. B. & Eriksson, S. Obstructive lung disease and trypsin inhibitors in alpha-1-antitrypsin deficiency. *Scandinavian Journal of Clinical and Laboratory Investigation* **19**, 205–208 (1967).
12. Fregonese, L. & Stolk, J. Hereditary alpha-1-antitrypsin deficiency and its clinical consequences. *Orphanet Journal of Rare Diseases* **3**, 16 (2008).
13. Calverley, P. M. A. & Koulouris, N. G. Flow limitation and dynamic hyperinflation: key concepts in modern respiratory physiology. *The European Respiratory Journal* **25**, 186–199 (2005).
14. Anderson, A. E. & Foraker, A. G. Centrilobular emphysema and panlobular emphysema: two different diseases 1. *Thorax* **28**, 547–550 (1973).
15. Kim, W. D. *et al.* Centrilobular and panlobular emphysema in smokers. Two distinct morphologic and functional entities. *The American Review of Respiratory Disease* **144**, 1385–1390 (1991).
16. Silverman, E. K. & Sandhaus, R. A. Clinical practice. Alpha1-antitrypsin deficiency. *The New England Journal of Medicine* **360**, 2749–2757 (2009).

17. Marciniuk, D., Hernandez, P. & Balter, M. Alpha-1 antitrypsin deficiency targeted testing and augmentation therapy: A Canadian Thoracic Society clinical practice guideline. *Canadian Thoracic* 109–116 (2012).
18. Abboud, R. T. & Vimalanathan, S. Pathogenesis of COPD. Part I. The role of protease-antiprotease imbalance in emphysema. *The international journal of tuberculosis and lung disease the official journal of the International Union against Tuberculosis and Lung Disease* **12**, 361–7 (2008).
19. Restrepo, G. L. & Heard, B. E. Mucous gland enlargement in chronic bronchitis: Extent of enlargement in the tracheo-bronchial tree. *Thorax* **18**, 334–339 (1963).
20. Dunnill, M. S., Massarella, G. R. & Anderson, J. A. A comparison of the quantitative anatomy of the bronchi in normal subjects, in status asthmaticus, in chronic bronchitis, and in emphysema. *Thorax* **24**, 176–179 (1969).
21. Jeffery, P. K. Remodeling in Asthma and Chronic Obstructive Lung Disease. *Critical Care Medicine* **164**, S28–38 (2001).
22. Silverman, E. K. Progress in Chronic Obstructive Pulmonary Disease Genetics. *Proceedings of the American Thoracic Society* **3**, 405–408 (2006).
23. Bossé, Y. Genetics of chronic obstructive pulmonary disease: a succinct review, future avenues and prospective clinical applications. *Pharmacogenomics* **10**, 655–667 (2009).
24. Bossé, Y. Updates on the COPD gene list. *International Journal of COPD* **7**, 607–631 (2012).
25. Hoffmann, D., Hoffmann, I. & El-Bayoumy, K. The less harmful cigarette: a controversial issue. a tribute to Ernst L. Wynder. *Chemical Research in Toxicology* **14**, 767–790 (2001).
26. Bossé, Y. *et al.* Molecular Signature of Smoking in Human Lung Tissues. *Cancer research* 1–11 (2012).
27. Martin, T. R. & Frevert, C. W. Innate immunity in the lungs. *Proceedings of the American Thoracic Society* **2**, 403–411 (2005).
28. Carnevali, S. *et al.* Cigarette smoke extract induces oxidative stress and apoptosis in human lung fibroblasts. *American Journal of Physiology Lung Cellular and Molecular Physiology* **284**, L955–L963 (2003).
29. MacNee, W. Oxidants/antioxidants and COPD. *Chest* **117**, 303S–17S (2000).
30. Sopori, M. Effects of cigarette smoke on the immune system. *Nature Reviews Immunology* **2**, 372–377 (2002).
31. Willemsse, B. W. M. *et al.* Effect of 1-year smoking cessation on airway inflammation in COPD and asymptomatic smokers. *The European Respiratory Journal* **26**, 835–45 (2005).
32. Greene, C. M., Hassan, T., Molloy, K. & McElvaney, N. G. The role of proteases, endoplasmic reticulum stress and SERPINA1 heterozygosity in

- lung disease and α -1 anti-trypsin deficiency. *Expert Review of Respiratory Medicine* **5**, 395–411 (2011).
33. Crystal, R. G. Alpha 1-antitrypsin deficiency, emphysema, and liver disease. Genetic basis and strategies for therapy. *Journal of Clinical Investigation* **85**, 1343–1352 (1990).
 34. Zimmerman, M. R., Jensen, A. M. & Sheehan, G. W. Agnaiyaaq: the autopsy of a frozen Thule mummy. *Arctic Anthropology* **37**, 52–59 (2000).
 35. Lacey, B., Sveger, T., Krams, A., Cernevska, G. & Krumina, A. Age of SERPINA1 gene PI Z mutation: Swedish and Latvian population analysis. *Annals of Human Genetics* **72**, 300–304 (2008).
 36. Abboud, R., Jung, B. & Mattman, A. Alpha1-antitrypsin deficiency: a clinical-genetic overview. *The Application of Clinical Genetics* **4**, 55–65 (2011).
 37. De Serres, F. J., Blanco, I. & Fernández-Bustillo, E. PI S and PI Z alpha-1 antitrypsin deficiency worldwide. A review of existing genetic epidemiological data. *Monaldi Archives for Chest Disease* **67**, 184–208 (2007).
 38. Ezzati, M. Indoor air pollution and health in developing countries. *Lancet* **366**, 104–106 (2005).
 39. Anthonisen, N. R. *et al.* The effects of a smoking cessation intervention on 14.5-year mortality. *Annals of Internal Medicine* **143**, 615; author reply 615 (2005).
 40. Praditsuwan, R., Assantachai, P., Wasi, C., Puthavatana, P. & Kositanont, U. The efficacy and cost effectiveness of vaccination against influenza among elderly persons living in the community. *Journal of the Medical Association of Thailand* **331**, 256–264 (2005).
 41. Vieira, D. S. R., Maltais, F. & Bourbeau, J. Home-based pulmonary rehabilitation in chronic obstructive pulmonary disease patients. *Current Opinion in Pulmonary Medicine* **16**, 134–43 (2010).
 42. Maltais, F. *et al.* Effects of Home-Based Pulmonary Rehabilitation in Patients with Chronic Obstructive Pulmonary Disease. *Annals of Internal Medicine* **149**, 869–878 (2008).
 43. 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. *Australian and New Zealand Journal of Medicine* **29**, 59–65 (1999).
 44. Stoller, J. K., Panos, R. J., Krachman, S., Doherty, D. E. & Make, B. Oxygen therapy for patients with COPD: current evidence and the long-term oxygen treatment trial. *Chest* **138**, 179–187 (2010).
 45. Burge, P. S. *et al.* Randomised, double blind, placebo controlled study of fluticasone propionate in patients with moderate to severe chronic obstructive pulmonary disease: the ISOLDE trial. *Bmj Clinical Research Ed.* **320**, 1297–1303 (2000).

46. Anthonisen, N. R. *et al.* Effects of smoking intervention and the use of an inhaled anticholinergic bronchodilator on the rate of decline of FEV1. *Quality* **272**, 1497 – 505 (1994).
47. Pauwels, R. A. *et al.* Long-term treatment with inhaled budesonide in persons with mild chronic obstructive pulmonary disease who continue smoking. *The New England Journal of Medicine* **340**, 1948–1953 (1999).
48. Vestbo, J. *et al.* Long-term effect of inhaled budesonide in mild and moderate chronic obstructive pulmonary disease: a randomised controlled trial. *Lancet* **353**, 1819–1823 (1999).
49. Project, G., Asia, E., Africa, S., Figs, S. & Tables, S. An integrated map of genetic variation from 1,092 human genomes. *Nature* **135**, 0–9 (2012).
50. Schadt, E. E. *et al.* Mapping the Genetic Architecture of Gene Expression in Human Liver. *PLoS Biology* **6**, 13 (2008).
51. Laurell, C. & Eriksson, S. The Electrophoretic α -1-Globulin Pattern of Serum in α -1-Antitrypsin Deficiency. *Scandinavian Journal of Clinical & Laboratory Investigation* **15**, 132–140 (1963).
52. Summary, E. American Thoracic Society/European Respiratory Society statement: standards for the diagnosis and management of individuals with alpha-1 antitrypsin deficiency. *American Journal of Respiratory and Critical Care Medicine* **168**, 818–900 (2003).
53. Machii, R. *et al.* Examination of the molecular diversity of alpha1 antitrypsin in urine: deficit of an alpha1 globulin fraction on cellulose acetate membrane electrophoresis. *Journal of clinical laboratory analysis* **19**, 16–21 (2005).
54. Borawski, J., Naumnik, B. & Myśliwiec, M. Serum α 1 -Antitrypsin But Not Complement C3 and C4 Predicts Chronic Inflammation in Hemodialysis Patients. *Renal Failure* **25**, 589–593 (2003).
55. Zhan, S. H. *et al.* Sanger sequencing solved a cryptic case of severe alpha-antitrypsin deficiency. *Clinical biochemistry* **45**, 499–501 (2012).
56. Hancock, D. B. *et al.* Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nature Genetics* **42**, 45–52 (2010).
57. Repapi, E. *et al.* Genome-wide association study identifies five loci associated with lung function. *Nature Genetics* **42**, 36–44 (2010).
58. Wilk, J. B. *et al.* A Genome-Wide Association Study of Pulmonary Function Measures in the Framingham Heart Study. *PLoS Genetics* **5**, 8 (2009).
59. Pillai, S. G. *et al.* A Genome-Wide Association Study in Chronic Obstructive Pulmonary Disease (COPD): Identification of Two Major Susceptibility Loci. *PLoS Genetics* **5**, 8 (2009).
60. Soler Artigas, M. *et al.* Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nature Genetics* **i**, 1082–90 (2011).

61. Wilk, J. B., Walter, R. E., Laramie, J. M., Gottlieb, D. J. & O'Connor, G. T. Framingham Heart Study genome-wide association: results for pulmonary function measures. *BMC Medical Genetics* **8**, S8 (2007).
62. Kong, X. *et al.* Genome-wide Association Study Identifies *BICD1* as a Susceptibility Gene for Emphysema. *American Journal of Respiratory and Critical Care Medicine* **183**, 43–49 (2011).
63. Cho, M. H. *et al.* A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Human molecular genetics* **21**, 947–57 (2012).
64. Cho, M. H. *et al.* Variants in *FAM13A* are associated with chronic obstructive pulmonary disease. *Nature Genetics* **42**, 200–202 (2010).
65. Chanock, S. J. *et al.* Replicating genotype-phenotype associations. *Nature* **447**, 655–60 (2007).
66. Guo, Y. *et al.* Genetic analysis of *IREB2*, *FAM13A* and *XRCC5* variants in Chinese Han patients with chronic obstructive pulmonary disease. *Biochemical and Biophysical Research Communications* **415**, 284–287 (2011).
67. Kingsley, C. B. Identification of Causal Sequence Variants of Disease in the Next Generation Sequencing Era. *Disease Gene Identification* **700**, 37–46 (2011).
68. Purcell, S. *et al.* PLINK: a toolset for whole genome association and population-based linkage analyses. *Genome Research* **81**, 559–75 (2007).
69. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–59 (2000).
70. Thorisson, G. & Smith, A. The international HapMap project web site. *Genome research* **15**, 1592–1593 (2005).
71. Thorgeirsson, T. E. *et al.* A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* **452**, 638–42 (2008).
72. Hung, R. J. *et al.* A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* **452**, 633–7 (2008).
73. Amos, C. I. *et al.* Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nature Genetics* **40**, 616–22 (2008).
74. Silverman, E. K. *et al.* Opportunities and Challenges in the Genetics of COPD 2010: An International COPD Genetics Conference Report. *COPD* **8**, 121–135 (2011).
75. Heinzen, E. L. *et al.* Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS biology* **6**, e1 (2008).
76. Qiu, W. *et al.* Genetics of sputum gene expression in chronic obstructive pulmonary disease. *PLoS one* **6**, e24395 (2011).

