ERJ Express. Published on May 12, 2016 as doi: 10.1183/13993003.00114-2016

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

To the Editor:

The major histocompatibility complex (MHC) has been linked with hundreds of diseases [1]. The MHC is one of the most complex regions of the human genome, because of the high gene density, extended linkage disequilibrium (LD) and sequence diversity [2]. Recent genome-wide association studies (GWAS) have identified polymorphisms located in the MHC that are associated with lung diseases and related traits: asthma, cystic fibrosis, idiopathic interstitial pneumonia, lung cancer and lung function. However, due to the limitations of GWAS and tissue-specific characteristics of gene expression [3], the causal genes and genetic mechanisms mediating the heritable risk within this locus remain to be found.

The present study has two goals: 1) to identify lung expression quantitative trait loci (eQTL) within the MHC region; and 2) to identify new susceptibility genes for lung diseases/traits by overlaying lung eQTL results and MHC single nucleotide polymorphisms (SNPs) previously associated with lung function and respiratory diseases. Susceptibility alleles for respiratory diseases that function as strong lung eQTL should facilitate the biological interpretation of GWAS results and the identification of causal genes in loci with high gene density and high LD such as the MHC.

Study subjects and lung specimens have been described previously [4]. Subjects were from three academic sites: Laval University (Quebec, Canada), University of British-Columbia (Vancouver, Canada) and Groningen University (Groningen, the Netherlands), henceforth referred to as Laval, UBC and Groningen, respectively. Genome-wide gene expression and genotyping profiles were obtained using a custom Affymetrix array (GPL10379; Affymetrix, Santa Clara, CA, USA) and the Illumina Human1M-Duo BeadChip array (Illumina, Inc., San Diego, CA, USA), respectively. Only subjects that passed genotyping and gene expression quality controls [5] were included in this study. Subjects with missing values for smoking status were also excluded, leaving 409, 287 and 342 subjects from Laval, UBC and Groningen, respectively.

The borders of the extended MHC (xMHC) were defined [1] and delimited by two genes on chromosome 6: *HIST1H2AA* and *KIFC1*. The expressions of all probe sets located within this region were analysed, which included 271 probe sets covering 212 transcripts. 6872 genotyped SNPs were available in the three datasets. Lung eQTL in the xMHC region were identified with expression data adjusted for age, sex and smoking status [5]. Association tests between adjusted expression traits and SNPs were performed using quantitative association tests implemented in PLINK (http://pngu.mgh.harvard.edu/purcell/plink/). Each possible combination of genotyped SNP and probe set located in the xMHC was tested in the Laval, UBC and Groningen datasets separately. eQTL were considered significant if they passed Bonferroni correction in Laval (0.05/(271 probe sets×6872 SNPs); p-values $\leqslant 2.68\times10^{-08}$) and were replicated (p-value $\leqslant 0.05$) in UBC and Groningen. Local- and distant-acting eQTL were defined using a 1 Mb threshold between the SNP and transcriptional start site positions.

A total of 5790 significant lung eQTL were detected in the xMHC. The eQTL consisted of 2807 SNPs (663 are independent, $r^2<0.8$) and 66 associated probe sets interrogating 50 transcripts. The majority of eQTL were local (n=5516), but 274 distant-acting eQTL were found. GWAS SNPs located in the xMHC and associated with lung diseases/traits were further investigated. The NHGRI-EBI (National Human Genome Research Institute-European Bioinformatics Institute) Catalog of published GWAS was downloaded (www.ebi.ac.uk/gwas/; date last accessed March 19, 2015) and hits located in the xMHC were isolated. Five lung diseases/traits were considered: asthma, cystic fibrosis lung disease severity, idiopathic interstitial pneumonia, lung cancer and lung function. Overall, a total of 27 risk associated genetic variants were considered, with GWAS p-values ranging from 4×10^{-6} and 4×10^{-23}. The eQTL significant threshold in Laval was set at a p-value of $\leqslant 6.83\times10^{-06}$ (0.05/(271 probe sets×27 SNPs)).

Among the 27 GWAS SNPs associated with lung diseases in the xMHC, 14 were associated with the expression levels of at least one transcript in Laval (figure 1a) and the same direction of effect was

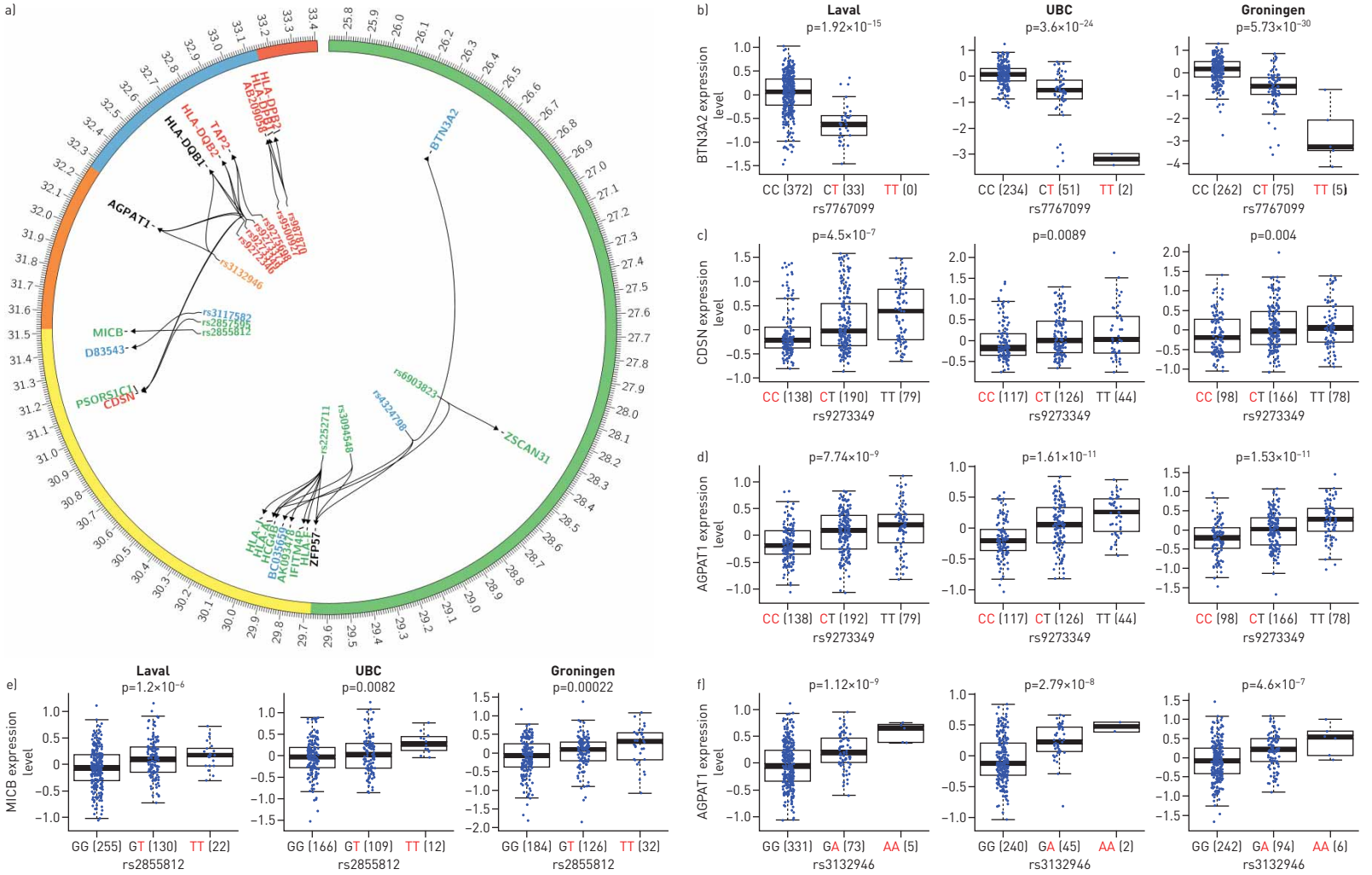

FIGURE 1 Lung expression quantitative trait loci (eQTL) in the extended major histocompatibility complex (xMHC) for genome-wide association study (GWAS)-nominated single nucleotide polymorphisms (SNPs) of lung diseases and traits. a) Graphical representation of significant lung eQTL identified in the xMHC considering only GWAS-associated SNPs for lung diseases (asthma, lung cancer and interstitial lung disease) and lung function. The outer border of the circle is the chromosomal position in Mb. xMHC sub-regions are colour coded (green: extended class I; yellow: classical class I; orange: classical class III; blue: classical class II; red: extended class II). SNPs and transcripts are colour coded according to their GWAS-associated trait or disease (green: pulmonary function; blue: lung cancer; orange: interstitial lung disease; red: asthma). Transcripts in black are associated with GWAS SNPs of two different traits or diseases. Black arrow heads represent the location of eQTL-genes. b–f) Boxplots of gene expression levels in the lung according to genotype groups. The y-axis shows the mRNA expression levels. The x-axis represents the three genotype groups for the SNP with the number of individuals in parenthesis. The risk allele identified in GWAS is shown in red. Box boundaries, whiskers and centre mark in boxplots represent the first and third quartiles, the most extreme data point which is no more than 1.5 times the interquartile range, and the median, respectively. The left, centre and right panels show the results for Laval, University of British Columbia (UBC) and Groningen samples, respectively. b) Lung cancer-associated SNP rs7767099 (*BTN3A2*). c) Asthma-associated SNP rs9273349 (*CDSN*). d) Asthma-associated SNP rs9273349 (*AGPAT1*). e) Lung function-associated SNP rs2855812 (*MICB*). f) Interstitial lung disease-associated SNP rs3132946 (*AGPAT1*).

observed in the two replication sets. 19 unique eQTL were found with SNPs previously associated with asthma (six SNPs; eight transcripts), two with idiopathic interstitial pneumonia (one SNP; two transcripts), four with lung cancer (two SNPs; four transcripts), 13 with lung function (five SNPs; 10 transcripts), and none with cystic fibrosis lung disease severity. The most relevant lung eQTL are illustrated in figure 1.

For lung cancer, rs7767099 was associated with mRNA expression of *BTN3A2* encoding the butyrophilin, subfamily 3, member A2. The surrogate risk-allele of rs7767099 in LD with the lung cancer-related SNP rs4324798 [6] was associated with lower mRNA expression levels of *BTN3A2* (figure 1b). *BTN3A2* was previously associated with ovarian cancer [7]. The same risk allele (rs7767099-T) was also associated with higher expression of *ZFP57* encoding the transcriptional repressor zinc finger protein 57 (eQTL p-values Laval 3.36×10^{-10}, UBC 5.03×10^{-17} and Groningen 4.95×10^{-18}). *ZFP57* is known to induce *IGF2* expression [8] and high protein expression of IGF2 is associated with a lower survival rate in lung cancer [9].

For asthma, the risk allele (rs9273349-C) was associated with lowered mRNA expression levels of *CDSN* (corneodesmosin) (figure 1c). This gene has been associated with skin diseases [10]. Genetic variants in this gene may alter the skin barrier contributing to allergic sensitisation and allergic asthma akin to the filaggrin gene [11]. The same risk variant was also associated with lowered expression levels of *AGPAT1* encoding 1-acylglycerol-3-phosphate O-acyltransferase 1 (figure 1d). AGPAT1 is an enzyme converting lysophosphatidic acid (LPA) into phosphatidic acid. The autotaxin–LPA pathway was recently associated with allergic lung inflammation [12].

The variant associated with lower lung function (rs2855812-T) was associated with higher mRNA expression of *MICB* encoding the MHC class I polypeptide-related sequence B (figure 1e). *MICB* encodes a ligand for the NKG2D type II receptor, which is a key mediator of cigarette smoke-stimulated natural killer cell hyperresponsiveness [13]. Higher levels of *MICB* may predispose individuals to greater inflammation due to the constant activation of the NKG2D receptor.

Finally, the idiopathic interstitial pneumonia risk allele (rs3132946-A) was associated with higher expression of *AGPAT1* (figure 1f). Accordingly, the risk variants for asthma (rs9273349) and idiopathic interstitial pneumonia (rs3132946) have an opposite direction of effect on the expression of *AGPAT1* in lung tissue (figures 1d and 1f). These two SNPs are in low LD in our datasets (max r^2=0.17). Elevated levels of LPA and/or autotaxin have been observed in bronchoalveolar lavage [14] and lung [15] of patients with idiopathic interstitial pneumonia. The risk-associated genetic variants for idiopathic interstitial pneumonia in the xMHC may thus be mediated through *AGPAT1*.

This study has limitations. The eQTL obtained in this study are derived from nontumour lung parenchymal samples. A proportion of eQTL are cell type- and tissue-specific [3]. Whole lung tissues contained multiple cell types and many eQTL in less prevalent cellular types will not be detected. eQTL are also known to be influenced by disease states. In this study, the majority of patients had lung cancer, but patients with other lung diseases were also included in the replication sets.

In conclusion, integration of lung specific eQTL located in the xMHC with results from previous GWAS for lung diseases/traits has revealed new susceptibility genes for lung cancer (*BTN3A2* and *ZFP57*), asthma (*AGPAT1* and *CDSN*), lung function (*MICB*) and idiopathic interstitial pneumonia (*AGPAT1*). This study provides significant new information concerning our understanding of the complex regulatory mechanism governing the xMHC region and sheds light on the mechanisms underlying GWAS loci for lung diseases identified in this region.

@ERSpublications
New susceptibility genes for asthma, idiopathic interstitial pneumonia, lung cancer and lung function in the MHC http://ow.ly/10onzl

Maxime Lamontagne[1], Philippe Joubert[1], Wim Timens[2], Dirkje S. Postma[3], Ke Hao[4], David Nickle[5], Don D. Sin[6,7], Peter D. Pare[6,7], Michel Laviolette[1] and Yohan Bossé[1,8]
[1]Institut universitaire de cardiologie et de pneumologie de Québec, Québec, QC, Canada. [2]University of Groningen, University Medical Center Groningen, Dept of Pathology and Medical Biology, GRIAC research institute, Groningen, The Netherlands. [3]University of Groningen, University Medical Center Groningen, Dept of Pulmonology, GRIAC Research Institute, Groningen, The Netherlands. [4]Dept of Genetics and Genomics Sciences, Mount Sinai School of Medicine, New York, NY, USA. [5]Merck & Co Inc, Rahway, NJ, USA. [6]University of British Columbia Center for Heart Lung Innovation and Institute for Heart and Lung Health, St Paul's Hospital, Vancouver, BC, Canada. [7]Respiratory Division, Dept of Medicine, University of British Columbia, Vancouver, BC, Canada. [8]Dept of Molecular Medicine, Laval University, Québec, QC, Canada.

Correspondence: Yohan Bossé, Laval University, Institut universitaire de cardiologie et de pneumologie de Québec, Pavillon Marguerite-d'Youville, Y2106, 2725, chemin Sainte-Foy, Québec (QC), Canada, G1V 4G5.
E-mail: yohan.bosse@criucpq.ulaval.ca

Received: Nov 09 2015 | Accepted after revision: March 29 2016

Support statement: The lung eQTL study at Laval University was supported by the Chaire de pneumologie de la Fondation JD Bégin de l'Université Laval, the Fondation de l'Institut universitaire de cardiologie et de pneumologie de Québec, the Respiratory Health Network of the Fonds de recherche Québec – Santé (FRQS), the Canadian Institutes of Health Research (MOP – 123369), and the Cancer Research Society and Read for the Cure. M. Lamontagne is the recipient of a doctoral studentship from the FRQS. D.D. Sin is a Tier 1 Canada Research Chair for COPD. Y. Bossé was the recipient of a Junior 2 Research Scholar from the FRQS and now holds a Canada Research Chair in Genomics of Heart and Lung Diseases. Funding information for this article has been deposited with FundRef.

Conflict of interest: Disclosures can be found alongside this article at erj.ersjournals.com

Acknowledgements: The authors would like to thank Christine Racine and Sabrina Biardel at the Institut universitaire de cardiologie et de pneumologie de Québec (Québec, QC, Canada) site of the Respiratory Health Network Biobank of the FRQS for their valuable assistance. At the University of British Columbia (Vancouver, BC, Canada) the authors thank the biobank coordinator Mark Elliott. At the University of Groningen, University Medical Center Groningen (Groningen, the Netherlands) Marnix Jonker is thanked for his support in selecting, handling and sending of lung tissues.

References

1 Horton R, Wilming L, Rand V, *et al.* Gene map of the extended human MHC. *Nat Rev Genet* 2004; 5: 889–899.
2 de Bakker PIW, Raychaudhuri S. Interrogating the major histocompatibility complex with high-throughput genomics. *Hum Mol Genet.* 2012; 21: 29–36.
3 Bossé Y. Genome-wide expression quantitative trait loci analysis in asthma. *Curr Opin Allergy Clin Immunol* 2013; 13: 487–494.
4 Hao K, Bossé Y, Nickle DC, *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* 2012; 8: e1003029.
5 Lamontagne M, Timens W, Hao K, *et al.* Genetic regulation of gene expression in the lung identifies CST3 and CD22 as potential causal genes for airflow obstruction. *Thorax* 2014; 69: 997–1004.
6 Landi MT, Chatterjee N, Yu K, *et al.* A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am J Hum Genet* 2009; 85: 679–691.
7 Le Page C, Marineau A, Bonza PK, *et al.* BTN3A2 expression in epithelial ovarian cancer is associated with higher tumor infiltrating T cells and a better prognosis. *PLoS One* 2012; 7: e38541.
8 Tada Y, Yamaguchi Y, Kinjo T, *et al.* The stem cell transcription factor ZFP57 induces IGF2 expression to promote anchorage-independent growth in cancer cells. *Oncogene* 2014; 34: 752–760.
9 Takanami I, Imamuma T, Hashizume T, *et al.* Insulin-like growth factor-II as a prognostic factor in pulmonary adenocarcinoma. *J Surg Oncol* 1996; 61: 205–208.
10 Orrù S, Giuressi E, Carcassi C, *et al.* Mapping of the major psoriasis-susceptibility locus (PSORS1) in a 70-Kb interval around the corneodesmosin gene (CDSN). *Am J Hum Genet* 2005; 76: 164–171.
11 Hudson TJ. Skin barrier function and allergic risk. *Nat Genet* 2006; 38: 399–400.
12 Park GY, Lee YG, Berdyshev E, *et al.* Autotaxin production of lysophosphatidic acid mediates allergic asthmatic inflammation. *Am J Respir Crit Care Med* 2013; 188: 928–940.
13 Wortham BW, Eppert BL, Motz GT, *et al.* NKG2D mediates NK cell hyperresponsiveness and influenza-induced pathologies in a mouse model of chronic obstructive pulmonary disease. *J Immunol* 2012; 188: 4468–4475.
14 Tager AM, LaCamera P, Shea BS, *et al.* The lysophosphatidic acid receptor LPA1 links pulmonary fibrosis to lung injury by mediating fibroblast recruitment and vascular leak. *Nat Med* 2008; 14: 45–54.
15 Oikonomou N, Mouratis MA, Tzouvelekis A, *et al.* Pulmonary autotaxin expression contributes to the pathogenesis of pulmonary fibrosis. *Am J Respir Cell Mol Biol* 2012; 47: 566–574.