Int. J. Dev. Biol. 62: 857-864 (2018)
https://doi.org/10.1387/ijdb.180210yb

THE INTERNATIONAL JOURNAL OF DEVELOPMENTAL BIOLOGY
www.intjdevbiol.com

Lung cancer susceptibility genetic variants modulate *HOXB2* expression in the lung

ALISSON CLEMENCEAU[#,1], OLIVIER BOUCHERAT[#,2], KIM LANDRY-TRUCHON[#,2], MAXIME LAMONTAGNE[1], SABRINA BIARDEL[1], PHILIPPE JOUBERT[1,3], STÉPHANE GOBEIL[4,5], BLANDINE SECCO[1], MATHIEU LAPLANTE[1,3], MATHIEU MORISSETTE[1,3], MA'EN OBEIDAT[7], WIM TIMENS[8], LUCIE JEANNOTTE*[,2,3,6] and YOHAN BOSSÉ*[,1,3,5]

[1]Institut Universitaire de Cardiologie et de Pneumologie de Québec, Université Laval, Québec, Canada, [2]CRCHU Québec-Université Laval, Québec, Canada, [3]Centre de Recherche sur le Cancer de l'Université Laval, Québec, Canada, [4]CRCHU de Québec, CHUL, Québec, Canada, [5]Department of Molecular Medicine, Université Laval, Québec, Canada, [6]Department of Molecular Biology, Medical Biochemistry and Pathology, Université Laval, Québec, Canada, [7]The University of British Columbia, Centre for Heart Lung Innovation, St Paul's Hospital, Vancouver, BC, Canada, [8]Department of Pathology and Medical Biology, University of Groningen, University Medical Center Groningen, GRIAC Research Institute, Groningen, The Netherlands

ABSTRACT The *HOX* genes are transcription factors that are expressed in coordinated spatio-temporal patterns to ensure normal development. Ectopic expression may instead lead to the development and progression of tumors. Genetic polymorphisms in the regions of four *HOX* gene clusters were tested for association with lung cancer in 420 cases and 3,151 controls. The effect of these variants on lung gene expression (expression quantitative trait loci, eQTL) was tested in a discovery set of 409 non-tumor lung samples and validated in two lung eQTL replication sets (n = 287 and 342). The expression levels of *HOXB2* were evaluated at the mRNA and protein levels by quantitative real-time PCR and immunohistochemistry in paired tumor and non-tumor lung tissue samples. The most significant SNP associated with lung cancer in the *HOXB* cluster was rs10853100 located upstream of the *HOXB* cluster. *HOXB2* was the top eQTL-regulated gene with several polymorphisms associated with its mRNA expression levels in lung tissue. This includes the lung cancer SNP rs10853100 that was significantly associated with *HOXB2* expression (P=3.39E-7). In the lung eQTL discovery and replication sets, the lung cancer risk allele (T) for rs10853100 was associated with lower *HOXB2* expression levels. In paired normal-tumor samples, *HOXB2* mRNA and protein levels were significantly reduced in tumors when compared to non-tumor lung tissues. Genetic variants in the *HOXB* cluster may confer susceptibility to lung cancer by modulating the expression of *HOXB2* in the lung.

KEY WORDS: *HOXB2, lung adenocarcinoma, lung eQTL*

Introduction

Aberrant expression of molecular regulators originally described as critical factors of mammalian embryogenesis may contribute to various pathologies, including cancer (Bonner *et al.*, 2004). Among the genes implicated in normal development, *HOX* genes act as master regulators (Bhatlekar *et al.*, 2014). They encode an evolutionarily conserved family of transcription factors, which regulate the expression of downstream effectors that direct pattern formation in animals. *HOX* genes specify embryonic regional identity and control morphogenesis (Krumlauf, 1994). In human and mouse, 39 *HOX* genes are ordered in four clusters located on different chromosomes (Table 1). Their position within a cluster correlates with their respective spatiotemporal expression during

Abbreviations used in this paper: eQTL, expression quantitative trait loci; FDR, false discovery rate; LD, linkage disequilibrium; NSCLC, non-small cell lung cancer; SNP, single nucleotide polymorphism.

***Address correspondence to:** Joint senior authors: Yohan Bossé. Institut universitaire de cardiologie et de pneumologie de Québec, Pavillon Marguerite-d'Youville, Y2106, 2725 chemin Sainte-Foy, Québec, QC, Canada, G1V 4G5. Tel: (418) 656-8711 ext. 3725. Fax: (418) 656-4940. E-mail: yohan.bosse@criucpq.ulaval.ca - https://orcid.org/0000-0002-3067-3711 or Lucie Jeannotte. Centre de recherche sur le cancer de l'Université Laval, CRCHU Québec-Université Laval, L'Hôtel-Dieu de Québec, 9, rue McMahon, Québec, QC, Canada G1R 3S3. Tel: (418) 525-4444 ext. 15552. Fax: (418) 691-5439. E-mail: lucie.jeannotte@crhdq.ulaval.ca - https://orcid.org/0000-0001-5578-420X - **#Note:** Joint first authors.

Supplementary Material (two tables) for this paper is available at: https://doi.org/10.1387/ijdb.180210yb

Submitted: 7 July, 2018; *Accepted:* 13 September, 2018.

ISSN: Online 1696-3547, Print 0214-6282

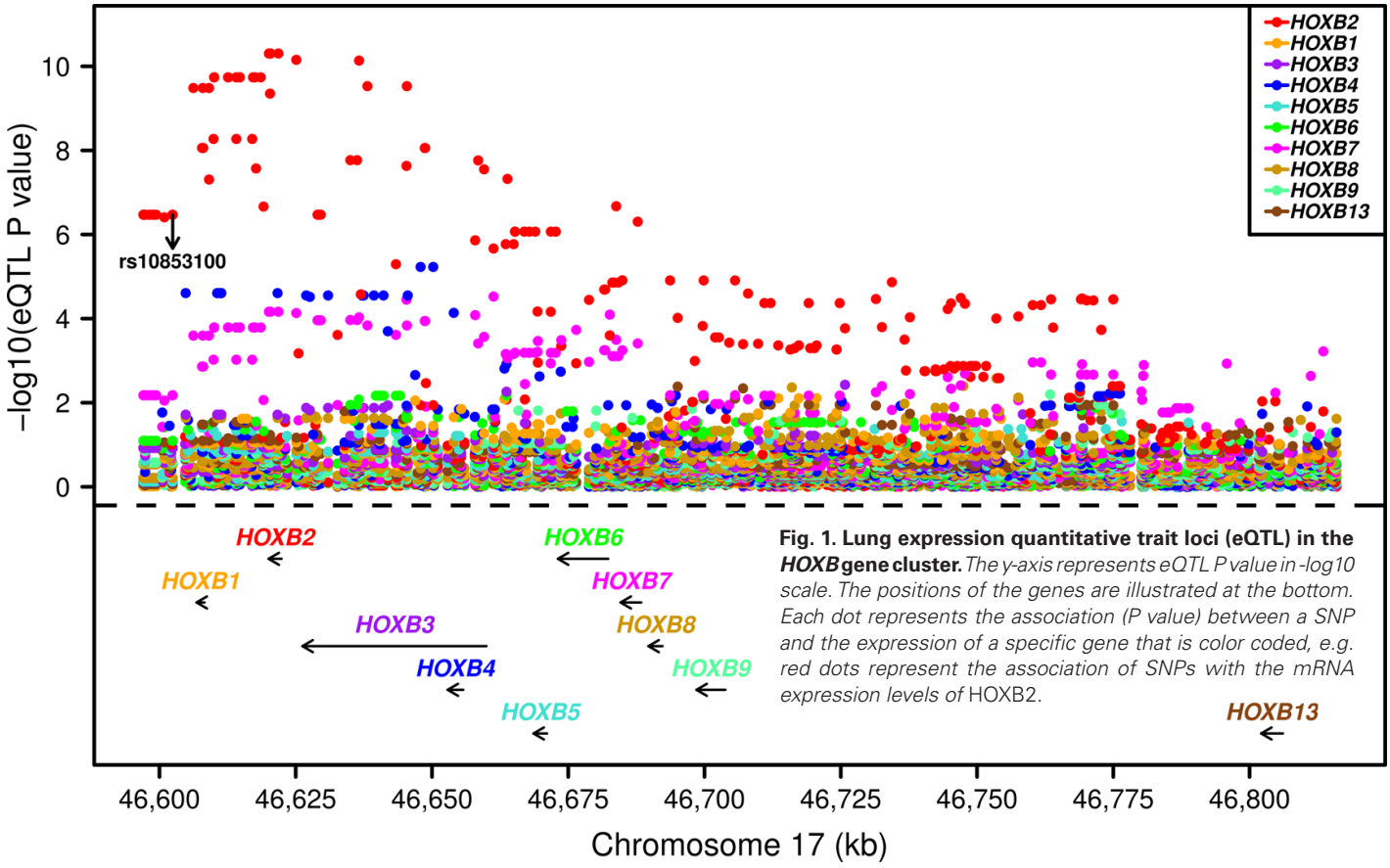


Fig. 1. Lung expression quantitative trait loci (eQTL) in the *HOXB* gene cluster. *The y-axis represents eQTL P value in -log10 scale. The positions of the genes are illustrated at the bottom. Each dot represents the association (P value) between a SNP and the expression of a specific gene that is color coded, e.g. red dots represent the association of SNPs with the mRNA expression levels of* HOXB2.

embryogenesis. This organization is essential for the specific regulation and function of each *HOX* gene and hence for the correct patterning of the embryo (Tschopp *et al.*, 2009). Accordingly, *HOX* genes are expressed in defined but overlapping domains along the developing axes.

In the lung, several *HOX* genes, mainly from the 3' end of the *HOXA* and *HOXB* clusters, are expressed along the respiratory tract, each with a distinct proximal-distal distribution (Herriges *et al.*, 2012). Mouse genetic studies have shown that only *Hoxa5* single mutant mice display respiratory tract defects, which cause respiratory distress and death at birth (Aubin *et al.*, 1997, Landry-Truchon *et al.*, 2017). This demonstrates the functional redundancy between *HOX* genes due to expression overlap (Boucherat *et al.*, 2013).

HOX genes are involved in multiple cellular processes, including cell differentiation, proliferation and survival, which once deregulated can lead to cancer (Abate-Shen, 2002). Inappropriate *HOX* gene expression is observed in multiple cancers supporting the notion that *HOX* deregulation is involved in the development and progression of tumors. For instance, *HOXC8* upregulation in human non-small cell lung cancer (NSCLC) correlates with tumor node metastasis and poor patient survival because gain of *HOXC8* expression promotes proliferation and migration via the activation of transforming growth factor beta 1 *(TGFB1)* expression (Liu *et al.*, 2018). The expression of *HOXB2* was also demonstrated as an adverse prognostic factor for stage I lung adenocarcinoma through its role in promoting invasion of lung cancer cells (Inamura *et al.*, 2008, Inamura *et al.*, 2007). In contrast, reduced *HOXA5* expression is reported to be an adverse prognosis indicator for NSCLC as *HOXA5* restrains NSCLC cell proliferation by activating the expression of cyclin-dependent kinase inhibitor p21 (Zhang *et al.*, 2015). However, other studies on NSCLC reported divergent data, and expression levels of *HOXA1*, *HOXA5*, *HOXA10* and *HOXC6* genes were found to be significantly increased in lung squamous cell carcinoma (Abe *et al.*, 2006). Higher expression of *HOXA5*, *HOXA10* and *HOXB2* was also reported in lung adenocarcinoma when compared to non-cancerous tissues (Abe *et al.*, 2006, Inamura *et al.*, 2007). These discrepancies make it challenging to decipher the role of *HOX* genes in lung cancer.

In the present study, we analyzed genetic polymorphisms in

TABLE 1

***HOX* GENES**

Gene clusters	Chromosome	Boundaries (bp)	Genes
HOXA	7	27,122,614-27,249,725	HOXA1, HOXA2, HOXA3, HOXA4, HOXA5, HOXA6, HOXA7, HOXA9, HOXA10, HOXA11, HOXA13
HOXB	17	46,596,807-46,816,111	HOXB1, HOXB2, HOXB3, HOXB4, HOXB5, HOXB6, HOXB7, HOXB8, HOXB9, HOXB13
HOXC	12	54,322,576-54,459,814	HOXC4, HOXC5, HOXC6, HOXC8, HOXC9, HOXC10, HOXC11, HOXC12, HOXC13
HOXD	2	176,954,530-177,065,635	HOXD1, HOXD3, HOXD4, HOXD8, HOXD9, HOXD10, HOXD11, HOXD12, HOXD13

TABLE 2

SNPs ASSOCIATED WITH LUNG CANCER WITHIN THE *HOX* GENE CLUSTERS

Cluster	SNP	Tested allele	Frequency cases	Frequency controls	P value
HOXA	rs10228276	G	0.20	0.17	0.0101
	rs13243033	T	0.20	0.17	0.0104
	rs6949451	T	0.52	0.47	0.0122
	rs3779456	C	0.38	0.43	0.0138
	rs6970537	A	0.38	0.43	0.0201
	rs2189239	T	0.10	0.08	0.0303
	rs2428432	T	0.29	0.25	0.0342
	rs10233387	A	0.49	0.45	0.0364
HOXB	rs10853100	T	0.54	0.49	0.0063
	rs11079833	T	0.04	0.03	0.0238
	rs733920	T	0.12	0.15	0.0394
	rs890431	T	0.07	0.09	0.0464
	rs15689	C	0.29	0.26	0.0479
HOXC	rs2067554	T	0.43	0.37	0.0031
	rs1386016	A	0.38	0.42	0.0232
HOXD	rs3754983	A	0.09	0.13	0.0050

TABLE 3

CLINICAL CHARACTERISTICS OF THE LUNG eQTL DISCOVERY AND REPLICATION SETS

	Discovery set (n = 409)	Replication 1 (UBC, n = 287)	Replication 2 (Groningen, n = 342)
Sex (male:female)	229:180	155:132	182:160
Age (years)*	63.3 ± 9.9	61.6 ± 13.2	51.5 ± 15.6
Smoking status, n (%)			
Never smokers	36 (9)	26 (9)	100 (29)
Former smokers	283 (69)	163 (57)	185 (54)
Current smokers	90 (22)	98 (34)	57 (17)
Primary diagnostic, n (%)			
Adenocarcinoma	235 (57)	78 (27)	36 (11)
Squamous cell carcinoma	104 (25)	82 (29)	51 (15)
Others**	70 (17)	127 (44)	255 (75)

*mean ± sd

**Others include patients that underwent lung surgery for other histological subtypes of lung cancer, COPD, alpha-1 antitrypsin deficiency, cystic fibrosis, idiopathic pulmonary fibrosis, bronchiectasis, pulmonary hypertension, and sarcoidosis.

TABLE 4

LUNG eQTLs DETECTED IN THE DISCOVERY SET FOR THE *HOX* CLUSTERS

HOX cluster	Gene	# of eQTL-SNP	Range P value	Range FDR
HOXA	HOXA9	1	1.49E-5	4.01E-3
HOXB	HOXB2	173	4.96E-11 to 1.15E-2	6.19E-9 to 3.81E-2
	HOXB4	23	5.90E-6 to 1.82E-3	9.54E-4 to 4.57E-2
	HOXB7	113	2.98E-5 to 8.87E-3	3.12E-3 to 4.52E-2
HOXC	HOXC6	42	4.85E-5 to 5.36E-3	9.45E-3 to 4.38E-2
HOXD	HOXD3	40	1.65E-5 to 6.55E-3	1.20E-3 to 3.75E-2
	HOXD1	25	7.30E-4 to 3.88E-3	1.33E-2 to 3.55E-2

Fig. 2. *HOXB2* gene expression levels in lungs according to genotype groups for SNP rs10853100 in the discovery set. *The y axis represents gene expression levels in the lung (n = 409). The x axis represents the three groups according to their genotypes. The number of subjects per groups is indicated in parentheses. The right y-axis shows the proportion of the gene expression variance explained by the SNP (black bar).*

the four *HOX* gene clusters for association with lung cancer. Gene expression levels of *HOX* genes were evaluated in non-tumor lung tissue and eQTL analyses were performed to identify genetic polymorphisms associated with mRNA expression levels of *HOX* genes in the lung. Genes regulated by lung cancer-associated polymorphisms were further characterized in paired tumor and non-tumor lung tissue samples at the mRNA and protein levels.

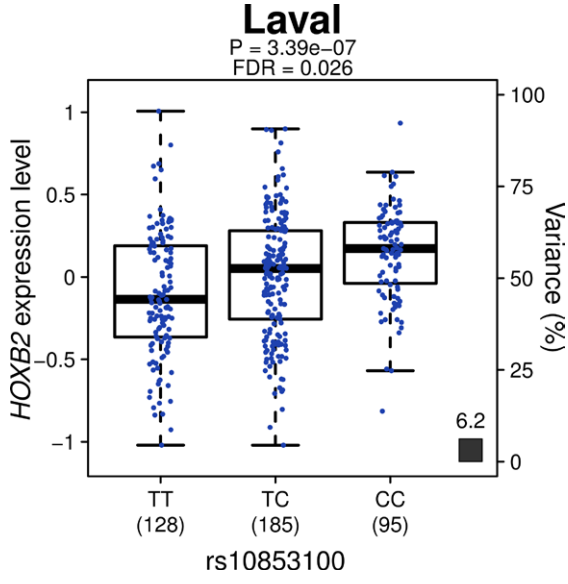
Results

Association between genetic polymorphisms in the HOX *gene clusters and lung cancer*

104 SNPs common to both cases (n = 420) and controls (n = 3,151) located within the *HOX* clusters were tested for association with lung cancer (Supplementary Table 1). SNPs passing a liberal P value of 0.05 are indicated in Table 2. The SNP most strongly associated with lung cancer, namely rs2067554, is located on human chromosome 12 in the *HOXC* cluster. This SNP is located near the 3'UTR of *HOXC4*, the most upstream *HOX* gene in this cluster. The second significant SNP associated with lung cancer in the cluster (rs1386016) is in modest LD with rs2067554 (r^2 = 0.37). A single SNP in the *HOXD* cluster on chromosome 2 was associated with lung cancer (rs3754983). This SNP is located at the 5' end of the *HOXD* cluster in the promoter region of *HOXD13*. Five SNPs were significantly associated with lung cancer in the *HOXB* cluster on chromosome 17. The most significant was rs10853100 located upstream of the *HOXB* cluster. Finally, 8 SNPs within the *HOXA* cluster on chromosome 7 were associated with lung cancer. These SNPs are located in four LD blocks: 1) rs2428432 located in the *HOXA2-HOXA3* intergenic region, 2) rs6949451, rs3779456 and rs6970537 located in the *HOXA10* gene, 3) rs2189239 and rs10233387 located within or near the *HOXA13* gene, and 4) rs10228276 and rs13243033 located in the most downstream region of the *HOXA* cluster.

Lung eQTLs in the HOX *gene clusters*

Table 3 shows the clinical characteristics of the final 409 subjects used in the eQTL analysis. Only one significant lung eQTL was



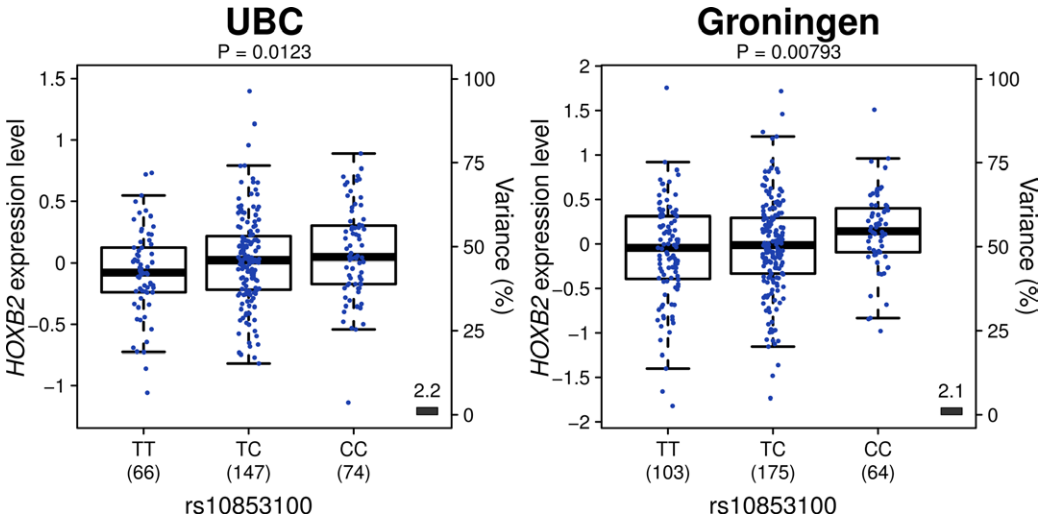


Fig. 3. Gene expression levels of the *HOXB2* gene in lungs according to genotype groups for SNP rs10853100 in the replication set from UBC (left panel) and the replication set from Groningen (right panel). *The* y *axis represents* HOXB2 *expression levels in the lung (n = 287 and 342 for UBC and Groningen, respectively). The* x *axis represents the three groups according to their genotypes. The number of subjects per groups is indicated in parentheses. The right y-axis shows the proportion of the gene expression variance explained by the SNP (black bar).*

found in the *HOXA* cluster at a FDR of 5% (Table 4). In contrast, the three other clusters were characterized by many lung eQTLs. Supplementary Table 2 provides the list of significant eQTLs in the four clusters. In the *HOXC* cluster, SNPs were associated with the expression of *HOXC6* (Table 4). For the *HOXD* cluster, SNPs were associated with the expression of *HOXD3* and *HOXD1* (Table 4). The highest number of eQTLs detected in this study were located in the *HOXB* complex (Table 4). The significant levels of eQTLs were also orders of magnitude higher in the *HOXB* cluster compared to the other clusters ($P=10^{-9}$ vs $P=10^{-3}$). We have thus focused subsequent analyses on the *HOXB* cluster. Fig. 1 shows the eQTL results of the *HOX* genes residing in this region. A total of 173 SNPs regulated the expression of *HOXB2*, 23 for *HOXB4*, and 113 for *HOXB7*. Many SNPs regulating the expression of *HOXB2* also regulated *HOXB7* expression, suggesting co-regulation of the two genes. However, the association signals were predominantly higher for *HOXB2*. Interestingly, the SNP most strongly associated with lung cancer in this region (rs10853100; Table 2) was significantly associated with the expression of *HOXB2* (Fig. 2). The lung cancer risk allele (T) for this polymorphism was associated with lower *HOXB2* RNA expression levels in lung (Fig. 2).

Two replications sets were used to confirm the lung eQTL rs10853100-*HOXB2*. Table 3 shows the clinical characteristics of each replication set. This eQTL was reproduced in both UBC and Groningen replication sets, suggesting that the risk allele in the *HOXB* cluster on chromosome 17 mediates its effect on lung cancer development through regulation of *HOXB2* in the lung (Fig. 3).

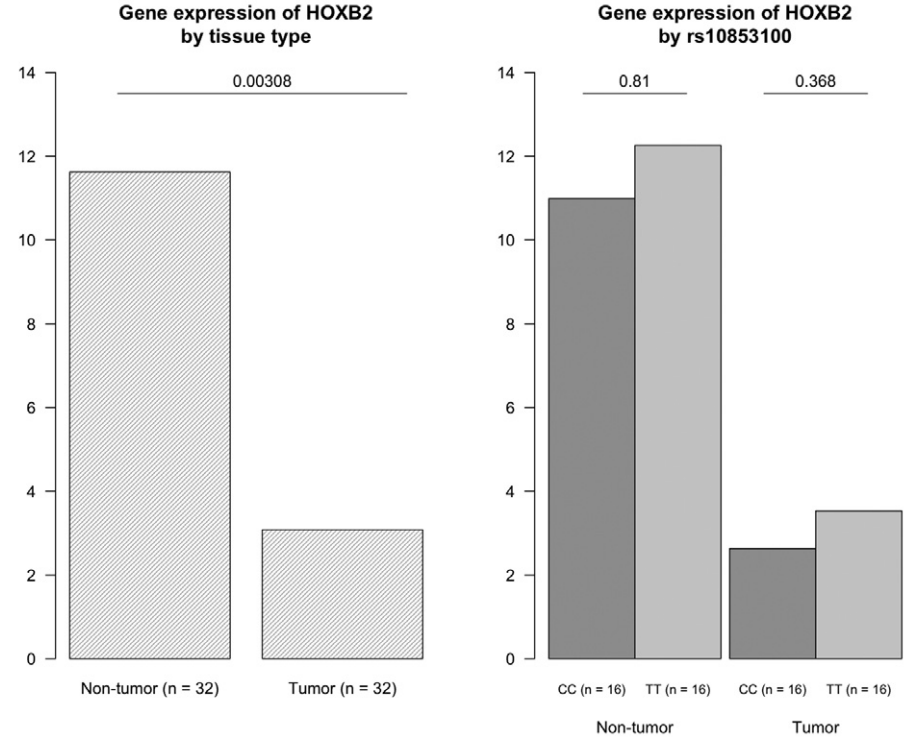


Fig. 4. *HOXB2* expression levels according to tissue types and genotype groups. *The y axis represents* HOXB2 *expression levels normalized to* GAPDH. *The number of individuals is indicated in parentheses at the bottom of each bar.*

Decreased HOXB2 *expression in lung tumor*

The association between rs10853100 and expression levels of *HOXB2* was further validated by qPCR analyses in tumor and non-tumor tissues from a different set of patients including 16 individuals with genotype CC and 16 with genotype TT. The clinical characteristics of patients by genotyping groups for the qPCR experiment are shown in Table 5. Fig. 4 shows *HOXB2* expression values grouped either by tissue types or genotype groups. *HOXB2* levels were significantly lowered in tumors compared to non-tumor lung tissues. In contrast, no difference was observed between the TT and CC genotypes in non-tumor tissues as well as in tumor specimens. Note that the pathological stage was not associated with *HOXB2* expression in these patients (Kruskal-Wallis chi-squared = 2.0638, P = 0.36).

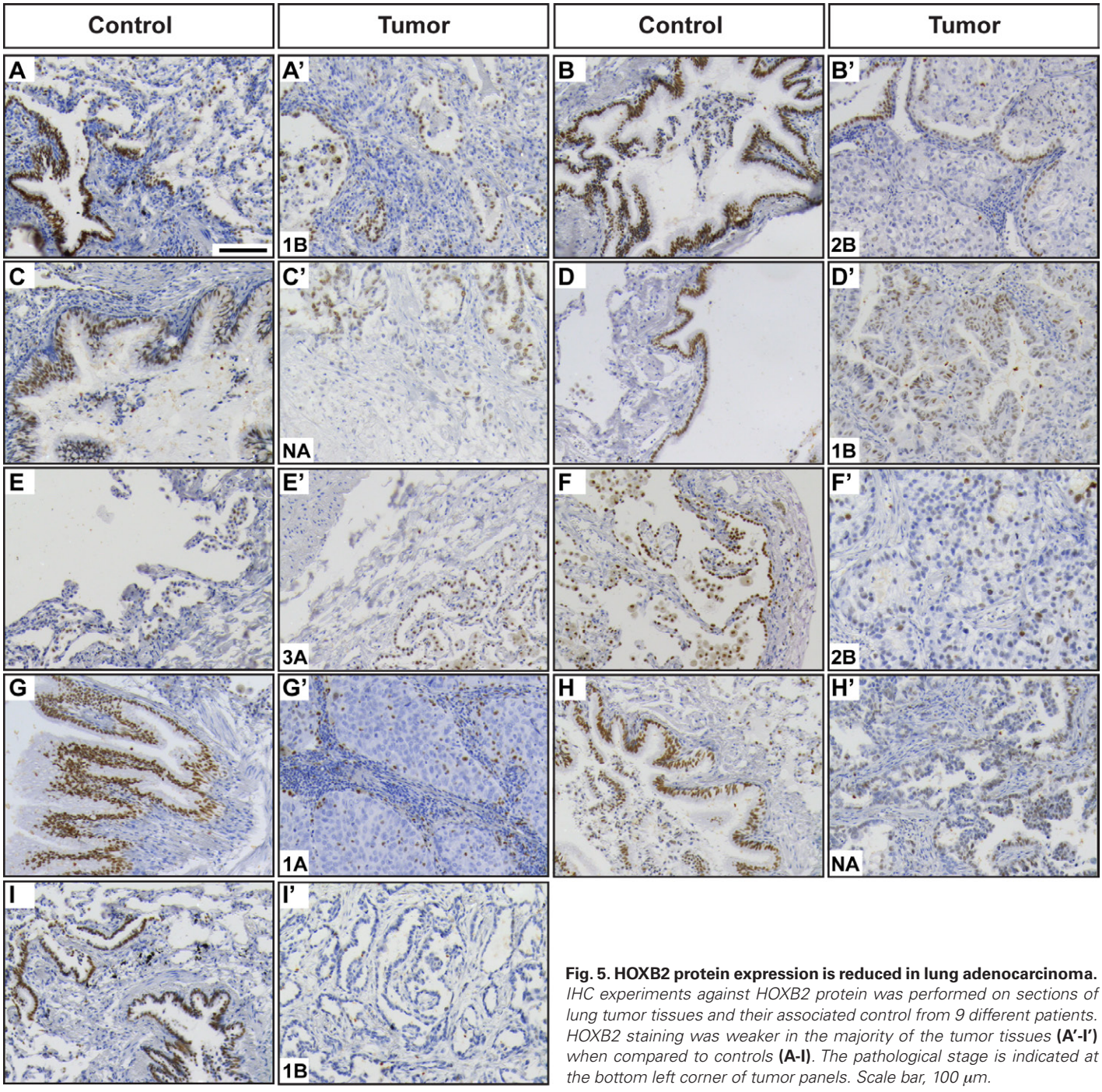


Fig. 5. HOXB2 protein expression is reduced in lung adenocarcinoma. *IHC experiments against HOXB2 protein was performed on sections of lung tumor tissues and their associated control from 9 different patients. HOXB2 staining was weaker in the majority of the tumor tissues* **(A'-I')** *when compared to controls* **(A-I)**. *The pathological stage is indicated at the bottom left corner of tumor panels. Scale bar, 100 μm.*

We also performed IHC experiments on lung adenocarcinoma and non-tumor specimens for HOXB2 protein expression. Consistent with the qPCR data, decreased HOXB2 protein expression was observed in the vast majority of the tumor specimens (Fig. 5).

Discussion

In this study, the variants in the *HOX* genes' family were investigated for their potential role in the susceptibility and pathogenesis of lung cancer, the leading cause of cancer death. Allele frequencies of germline variants located in the four *HOX* gene clusters were compared between patients and controls. The strongest evidences of genetic associations identified were for SNP rs2067554 located near the 3'UTR of *HOXC4* and SNP rs10853100 located upstream of the *HOXB* cluster. Using a large-scale lung eQTL dataset, *HOXB2* was found as the most statistically significant eQTL-regulated gene. Interestingly, the disease-associated SNP rs10853100 was associated with the expression of *HOXB2* and direction of effect was that the lung cancer risk allele (T) correlated with lowered expression of *HOXB2*. This direction of effect was confirmed in two large and independent lung eQTL datasets. The mRNA expression and protein levels of *HOXB2* were found to be reduced in tumors

TABLE 5

CLINICAL CHARACTERISTICS OF PATIENTS FOR THE VALIDATION STEP BY qPCR ACCORDING TO GENOTYPING GROUPS FOR SNP RS10853100

	CC	TT
Gender (male:female)	7:9	8:8
Age (years)*	63.6 ± 8.8	59.8 ± 11.2
Smoking status, n (%)		
Never smokers	1 (6)	0 (0)
Former smokers	13 (81)	9 (56)
Current smokers	2 (13)	7 (44)
Primary diagnostic, n (%)		
Adenocarcinoma	16 (100)	16 (100)
Squamous cell carcinoma	0 (0)	0 (0)
Others	0 (0)	0 (0)

*mean ± sd

compared to paired non-tumor lung tissues. Taken together, these results suggest that common variants in the *HOXB* cluster confer susceptibility to lung cancer through down-regulation of the *HOXB2* mRNA expression levels in the lung.

The mRNA expression level of *HOXB2* has been identified as a prognostic indicator for stage I lung adenocarcinomas in a Japanese population (Inamura *et al.*, 2007). The group with higher expression (n=15) had worse disease-specific survival compared to the group with lower expression (n=37). In a subset of these cases with normal lung tissue samples available, *HOXB2* expression was higher in tumor relative to normal tissues. In the same study, *HOXB2* expression in normal lung tissues was higher in metastatic cancer patients. A follow-up study from the same group proposed that *HOXB2* promotes NSCLC cells invasion through the up-regulation of metastasis-related genes, including genes coding for glycoproteins and signaling molecules (Inamura *et al.*, 2008). Together these results suggested that higher *HOXB2* expression favor tumorigenesis and/or invasion. In contrast, in the current study, we demonstrated reduced mRNA and protein expression of *HOXB2* in adenocarcinomas compared to adjacent normal lung tissues. Although this seems contradictory, similar observations were made for a positive regulator of HOX, namely *PREP1* (Penkov *et al.*, 2013). *PREP1* was shown to be absent or down-regulated in human tumors compared to normal tissues (Longobardi *et al.*, 2010), as we observed for *HOXB2*. On the other hand, it was found to be a negative prognostic factor in patients with stage I NSCLC (Risolino *et al.*, 2014), as observed for *HOXB2* in the Japanese population described above. *PREP1* is recognized as a tumor suppressor that inhibits tumor growth, but at the same time an inducer of the epithelial-mesenchymal transition and a prometastatic transcription factor (Risolino *et al.*, 2014). It is thus possible that both up- and down-regulation are detrimental and a fine balance in *PREP1* and also *HOXB2* expression is required. In our recent transcriptomic study (Bossé *et al.*, 2017), *HOXB2* was not differentially expressed in stage I lung adenocarcinomas compared to adjacent non-tumor lung samples collected at 0, 2, 4, and 6 cm away from the tumor. In addition, *HOXB2* expression levels in non-tumor sites were relatively stable across the different distances from the tumor. Larger sample size studies will be needed to determine whether *HOXB2* is up- or down-regulated in lung tumor.

Changes in *HOXB2* expression were previously reported in several types of cancer. *HOXB2* was more expressed in pancreatic and cervical cancers (Gonzalez-Herrera *et al.*, 2015, Segara *et al.*, 2005). In contrast, expression was decreased in breast cancer cell lines and acute myeloid leukemia, suggesting that it can act as a tumor suppressor (Boimel *et al.*, 2011, Lindblad *et al.*, 2015). This supports the notion that aberrant expression of *HOXB2* gene is detrimental to normal cell growth.

In the qPCR experiment, no difference in *HOXB2* expression was observed between the TT and CC genotypes in non-tumor tissues as well as in tumor specimens. Lower expression was expected for TT, at least in non-tumor tissues, considering the results of the three independent eQTL datasets (n=409 Laval, 287 UBC, and 342 Groningen). We believe that the sample size was not sufficient in the qPCR experiment (n=32 patients) to demonstrate the effect of the genotype. However, this sample size was enough to demonstrate lower expression of *HOXB2* in tumors compared to non-tumor lung tissues.

Our study has limitations. First, the case-control genetic association study was conducted using a shared control group not phenotyped for lung cancer. Accordingly, some of the controls are likely to develop lung cancer in the future and may result is some misclassification and dilution of the signal. Second, the genetic association results were not replicated in an independent dataset. Instead, lung eQTL datasets were used to mechanically explain the genetic association findings into gene expression in the lung. Third, no functional work (*in vitro* and/or *in vivo*) was performed to demonstrate the causal role of *HOXB2* and disease-associated SNPs in lung cancer. Finally, the genetic association study and lung eQTL datasets were obtained from patients of European ancestry only. The results may thus not translate into other ethnic groups.

In summary, genetic variants in the *HOX* gene clusters show association with lung cancer. One of the top disease-associated variant was also associated with the expression levels of *HOXB2* in non-tumor lung tissues. The lung cancer risk allele (T) was associated with lower expression of *HOXB2* in three independent datasets (discovery and two replication sets). Consistent with this direction of effect, the mRNA and protein levels of *HOXB2* were lower in tumors compared to adjacent normal lung tissues. These results highlight *HOXB2* as a candidate gene for lung cancer and suggest that down-regulation of this gene as a potential mechanism involved in lung cancer development. Validation of these findings and functional studies are needed to establish the causality of *HOXB2* for lung cancer.

Materials and Methods

Discovery set

Lung specimens (n = 420) from lung cancer patients were obtained from the biobank of the Institut universitaire de cardiologie et de pneumologie de Québec (IUCPQ). Corresponding clinical data were obtained at the time of tissue collection and available in a local database including demographic variables and detailed pathology report. Written informed consent was obtained from all subjects. The study was approved by the local ethics committee.

Replication sets

Significant lung eQTL identified in the discovery set were replicated in two independent cohorts. The first replication set consists of patients recruited at the University of British Columbia (UBC; Vancouver, Canada). At UBC, the majority of samples were from patients undergoing resection of small peripheral lung lesions. Additional samples were from autopsy

and at the time of lung transplantation. Patients from UBC provided written informed consent and the study was approved by the ethics committee at the UBC-Providence Health Care Research Institute Ethics Board. The second replication set was recruited at the University of Groningen (Groningen, The Netherlands). At Groningen, the lung specimens were obtained at surgery from patients with various lung diseases, including patients undergoing therapeutic resection for lung tumors, harvested from a site distant from the tumor, and lung transplantation. 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). The final analyses were performed with 287 and 342 lung specimens from UBC and Groningen, respectively.

Gene expression and genotyping microarrays

Gene expression profiling and genotyping were described previously (Hao *et al*., 2012, Lamontagne *et al*., 2013). Briefly a non-neoplastic pulmonary parenchyma sample was harvested at surgery. RNA was extracted from the lung specimens and then hybridized to a custom Affymetrix array (GEO platform GPL10379). Expression values were extracted using the Robust Multichip Average (RMA) method (Irizarry *et al*., 2003) and probe sets testing *HOX* genes were considered in this study. Genotyping was carried using the Illumina Human1M-Duo BeadChip. Single nucleotide polymorphism located 10 kb up- and downstream of the four *HOX* clusters were analyzed in this study (Table 1).

Quantitative real-time PCR

A subset of patients with lung adenocarcinoma enrolled in the lung eQTL mapping study was selected for validation by quantitative real-time PCR (qPCR). They were selected based on tissue availability and using a random stratified strategy based on genotypes for the single nucleotide polymorphism (SNP) rs10853100. Tumor and non-tumor lung specimens from sixteen individuals with genotype CC and sixteen with genotype TT were evaluated. RNA from tumor and non-tumor specimens was extracted from 30 mg of frozen tissues using the RNeasy Plus Universal Mini kit (Qiagen, MD, USA). RNA concentration and purity were assayed by UV 260/280 nm ratio on a NanoVue spectrophotometer (GE Healthcare, NJ, USA). The quality of each RNA sample was then evaluated by automated electrophoresis on the Experion system (Bio-Rad, CA, USA). cDNA was synthesized with the Superscript II Reverse Transcriptase (Invitrogen) using random primers. Quantitative PCR (qPCR) was performed with Power SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA USA) and a thermal cycler ABI PRISM 7000. Samples were analyzed in triplicate. The primer sequences used for the *HOXB2* gene were: forward-5'TCCTTCCCGACCTCAACTTC3', reverse-5'GAGAACCCTGTAGGCTAGGG3'. The *GAPDH* was used as reference gene using the primers: forward-5'GGTCGGAGTCAACGGATTTG3' and reverse-5'GGAAGATGGTGATGGGATTTC3'.

Immunohistochemistry analysis

Lung adenocarcinoma and non-tumor tissues obtained from surgeries were fixed in 10% formaldehyde, embedded in paraffin and sectioned at 4 μm. Immunohistochemistry (IHC) assay was performed as described (Gendronneau *et al*., 2012). The primary HOXB2 antibody was a mouse monoclonal antibody against the human HOXB2 protein (1/100 dilution, Developmental Studies Hybridoma Bank). The secondary antibody was a goat anti-mouse biotinylated antibody (1/250, Jackson Lab). Slides were counterstained with hematoxylin.

Statistical analyses

Genetic association

Genotypes from the lung cancer cases from the discovery set (n=420) were compared to 3,151 publicly available controls taken from the Illumina® iControlDB. Genetic association tests comparing allele frequencies between cases and controls were performed with PLINK (Purcell *et al*., 2007).

Lung eQTL analyses

RNA expression traits 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 PLINK (Purcell *et al*., 2007). Significant eQTLs were those passing the 5% false discovery rate (FDR) threshold (Benjamini and Hochberg, 1995) in the discovery set. P values lower than 0.05 were considered statistically significant in the replication sets.

qPCR

Expression values (½ΔΔCT) deviating by more than three standard errors from the mean were considered outliers and removed from analysis. Gene expression values for *HOXB2* were analyzed by tissue types (non-tumor vs tumor) and genotypes (CC vs TT). Two-sided t-tests and paired t-tests were used to compare genotyping groups and tissue types, respectively.

Acknowledgments

The authors would like to thank the research staff at the Quebec Respiratory Health Network Biobank, site IUCPQ for their valuable assistance. This study was supported by the Fondation de l'Institut universitaire de cardiologie et de pneumologie de Québec (to Y.B.), the Quebec Respiratory Health Network (to Y.B.), Centre de Recherche sur le Cancer de l'Université Laval (to L.J. and Y.B.), Canadian Institutes of Health Research (MOP-15139 to L.J. and MOP-123369 to Y.B.), and Cancer Research Society (to L.J. and to Y.B.). Y.B. holds a Canada Research Chair in Genomics of Heart and Lung Diseases.

References

ABATE-SHEN C. (2002). Deregulated homeobox gene expression in cancer: cause or consequence? *Nat Rev Cancer* 2: 777-785.

ABE M, HAMADA J, TAKAHASHI O, TAKAHASHI Y, TADA M, MIYAMOTO M, MORIKAWA T, KONDO S and MORIUCHI T. (2006). Disordered expression of HOX genes in human non-small cell lung cancer. *Oncol Rep* 15: 797-802.

AUBIN J, LEMIEUX M, TREMBLAY M, BÉRARD J and JEANNOTTE L. (1997). Early postnatal lethality in Hoxa-5 mutant mice is attributable to respiratory tract defects. *Dev Biol* 192: 432-445.

BENJAMINI Y and HOCHBERG Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc Ser B* 57: 289-300.

BHATLEKAR S, FIELDS J Z and BOMAN B M. (2014). HOX genes and their role in the development of human cancers. *J Mol Med (Berl)* 92: 811-823.

BOIMEL P J, CRUZ C and SEGALL J E. (2011). A functional *in vivo* screen for regulators of tumor progression identifies HOXB2 as a regulator of tumor growth in breast cancer. *Genomics* 98: 164-172.

BONNER A E, LEMON W J, DEVEREUX T R, LUBET R A and YOU M. (2004). Molecular profiling of mouse lung tumors: association with tumor progression, lung development, and human lung adenocarcinomas. *Oncogene* 23: 1166-1176.

BOSSÉ Y, SAZONOVA O, GAUDREAULT N, BASTIEN N, CONTI M, PAGE S, TRAHAN S, COUTURE C and JOUBERT P. (2017). Transcriptomic microenvironment of lung adenocarcinoma. *Cancer Epidemiol Biomarkers Prev* 26: 389-396.

BOUCHERAT O, MONTARON S, BERUBE-SIMARD F A, AUBIN J, PHILIPPIDOU P, WELLIK D M, DASEN J S and JEANNOTTE L. (2013). Partial functional redundancy between Hoxa5 and Hoxb5 paralog genes during lung morphogenesis. *Am J Physiol Lung Cell Mol Physiol* 304: L817-830.

GENDRONNEAU G, BOUCHERAT O, AUBIN J, LEMIEUX M and JEANNOTTE L. (2012). The loss of Hoxa5 function causes estrous acyclicity and ovarian epithelial inclusion cysts. *Endocrinology* 153: 1484-1497.

GONZALEZ-HERRERA A, SALGADO-BERNABE M, VELAZQUEZ-VELAZQUEZ C, SALCEDO-VARGAS M, ANDRADE-MANZANO A, AVILA-MORENO F and PINA-SANCHEZ P. (2015). Increased expression of HOXB2 and HOXB13 proteins is associated with HPV infection and cervical cancer progression. *Asian Pac J Cancer Prev* 16: 1349-1353.

HAO K, BOSSÉ Y, NICKLE D C, PARE P D, POSTMA D S, LAVIOLETTE M, SANDFORD A, HACKETT T L, DALEY D, HOGG J C *et al.,* (2012). Lung eQTLs to Help Reveal the Molecular Underpinnings of Asthma. *PLoS Genet* 8: e1003029.

HERRIGES J C, YI L, HINES E A, HARVEY J F, XU G, GRAY P A, MA Q and SUN X. (2012). Genome-scale study of transcription factor expression in the branching mouse lung. *Dev Dyn* 241: 1432-1453.

INAMURA K, TOGASHI Y, NINOMIYA H, SHIMOJI T, NODA T and ISHIKAWA Y. (2008). HOXB2, an adverse prognostic indicator for stage I lung adenocarcinomas, promotes invasion by transcriptional regulation of metastasis-related genes in HOP-62 non-small cell lung cancer cells. *Anticancer Res* 28: 2121-2127.

INAMURA K, TOGASHI Y, OKUI M, NINOMIYA H, HIRAMATSU M, SATOH Y, OKUMURA S, NAKAGAWA K, SHIMOJI T, NODA T *et al.,* (2007). HOXB2 as a novel prognostic indicator for stage I lung adenocarcinomas. *J Thorac Oncol* 2: 802-807.

IRIZARRY R A, HOBBS B, COLLIN F, BEAZER-BARCLAY Y D, ANTONELLIS K J, SCHERF U and SPEED T P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4: 249-264.

KRUMLAUF R. (1994). Hox genes in vertebrate development. *Cell* 78: 191-201.

LAMONTAGNE M, COUTURE C, POSTMA D S, TIMENS W, SIN D D, PARÉ P D, HOGG J C, NICKLE D, LAVIOLETTE M and BOSSÉ Y. (2013). Refining susceptibility loci of chronic obstructive pulmonary disease with lung eQTLs. *PLoS One* 8: e70220.

LANDRY-TRUCHON K, HOUDE N, BOUCHERAT O, JONCAS F H, DASEN J S, PHILIPPIDOU P, MANSFIELD J H and JEANNOTTE L. (2017). HOXA5 plays tissue-specific roles in the developing respiratory system. *Development* 144: 3547-3561.

LINDBLAD O, CHOUGULE R A, MOHARRAM S A, KABIR N N, SUN J, KAZI J U and RONNSTRAND L. (2015). The role of HOXB2 and HOXB3 in acute myeloid leukemia. *Biochem Biophys Res Commun* 467: 742-747.

LIU H, ZHANG M, XU S, ZHANG J, ZOU J, YANG C, ZHANG Y, GONG C, KAI Y and LI Y. (2018). HOXC8 promotes proliferation and migration through transcriptional up-regulation of TGFbeta1 in non-small cell lung cancer. *Oncogenesis* 7: 1.

LONGOBARDI E, IOTTI G, DI ROSA P, MEJETTA S, BIANCHI F, FERNANDEZ-DIAZ L C, MICALI N, NUCIFORO P, LENTI E, PONZONI M *et al.,* (2010). Prep1 (pKnox1)-deficiency leads to spontaneous tumor development in mice and accelerates EmuMyc lymphomagenesis: a tumor suppressor role for Prep1. *Mol Oncol* 4: 126-134.

PENKOV D, MATEOS SAN MARTIN D, FERNANDEZ-DIAZ L C, ROSSELLO C A, TORROJA C, SANCHEZ-CABO F, WARNATZ H J, SULTAN M, YASPO M L, GABRIELI A *et al.,* (2013). Analysis of the DNA-binding profile and function of TALE homeoproteins reveals their specialization and specific interactions with Hox genes/proteins. *Cell Rep* 3: 1321-1333.

PURCELL S, NEALE B, TODD-BROWN K, THOMAS L, FERREIRA M A, BENDER D, MALLER J, SKLAR P, DE BAKKER P I, DALY M J *et al.,* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559-575.

RISOLINO M, MANDIA N, IAVARONE F, DARDAEI L, LONGOBARDI E, FERNANDEZ S, TALOTTA F, BIANCHI F, PISATI F, SPAGGIARI L *et al.,* (2014). Transcription factor PREP1 induces EMT and metastasis by controlling the TGF-beta-SMAD3 pathway in non-small cell lung adenocarcinoma. *Proc Natl Acad Sci USA* 111: E3775-3784.

SEGARA D, BIANKIN A V, KENCH J G, LANGUSCH C C, DAWSON A C, SKALICKY D A, GOTLEY D C, COLEMAN M J, SUTHERLAND R L and HENSHALL S M. (2005). Expression of HOXB2, a retinoic acid signaling target in pancreatic cancer and pancreatic intraepithelial neoplasia. *Clin Cancer Res* 11: 3587-3596.

TSCHOPP P, TARCHINI B, SPITZ F, ZAKANY J and DUBOULE D. (2009). Uncoupling time and space in the collinear regulation of Hox genes. *PLoS Genet* 5: e1000398.

ZHANG M L, NIE F Q, SUN M, XIA R, XIE M, LU K H and LI W. (2015). HOXA5 indicates poor prognosis and suppresses cell proliferation by regulating p21 expression in non small cell lung cancer. *Tumour Biol* 36: 3521-3531.