Human Heredity

Research Article

Hum Hered 2024;89:71–83
DOI: 10.1159/000539520

Received: January 3, 2024
Accepted: May 17, 2024
Published online: June 14, 2024

A Statistical Testing Strategy Accounting for Random and Nonrandom (Skewed) X-Chromosome Inactivation Identifies Lung Cancer Susceptibility Loci among Smokers

Rodolphe Jantzen[a] Sophie Camilleri-Broët[b] Nicole Ezer[b] Philippe Broët[c, d, e]

[a]CHU Sainte-Justine Research Center, Montréal, QC, Canada; [b]McGill University Health Centre, McGill University, Montréal, QC, Canada; [c]Ecole de Santé Publique de l'Université de Montréal (ESPUM), Université de Montréal, Montréal, QC, Canada; [d]Université Paris-Saclay, UVSQ, Université Paris-Sud, Inserm, High-Dimensional Biostatistics for Drug Safety and Genomics, CESP, Villejuif, France; [e]Assistance Publique-Hôpitaux de Paris (AP-HP), Hôpital Paul Brousse, Villejuif, France

Keywords
Lung cancer · X chromosome · Skewed X-chromosome inactivation · Additive-multiplicative hazard model

Abstract
Introduction: Lung cancer is the most common cancer worldwide in mortality and the second in incidence. Epidemiological studies found a higher lung cancer risk for smoking women in comparison to men, but these sex differences, irrespective of smoking habits, remain controversial. One of the hypotheses concerns the genetic contribution of the sex chromosomes. However, while genome-wide association studies identified many lung cancer susceptibility loci, these analyses have excluded X-linked loci. ***Methods:*** To account for nongenetic factors, we first presented an association test based on an additive-multiplicative hazard model accounting for random/nonrandom X-inactivation process. A simulation study was performed to investigate the properties of the proposed test as compared with the Wald test from a Cox model with random X-inactivation process and the partial likelihood ratio test proposed by Xu et al. accounting for nonrandom X-inactivation process. Then, we performed an X chromosome-wide association study on 9,261 individuals from the population-based cohort CARTaGENE to identify susceptibility loci for lung cancer among current and past smokers. We adjusted for the PLCOm2012 lung cancer risk score used in screening programs. ***Results:*** Simulation results show the good behavior of the proposed test in terms of power and type I error probability as compared to the Xu et al. and the Wald test. Using the proposed test statistic and adjusting for the PLCOm2012 score, the X chromosome-wide statistical analysis identified two SNPs in low-linkage disequilibrium located in the IL1RAPL1 (IL-1 R accessory protein-like) gene: rs12558491 ($p = 2.75\times10^{-9}$) and rs12835699 ($p = 1.26\times10^{-6}$). For both SNPs, the minor allele was associated with lower lung cancer risk. Adjusting for multiple testing, no signal was detected using the Wald or the Xu et al. likelihood ratio tests. ***Conclusion:*** By taking into account smoking behavior and the X-inactivation process, the investigation of the X chromosome has shed a new light on the association between X-linked loci and lung cancer. We identified two loci associated with lung cancer located in the IL1RAPL1 gene. This finding would have been overlooked by examining only results from other test statistics.

karger@karger.com
www.karger.com/hhe

Correspondence to:
Philippe Broët, philippe.broet@inserm.fr

Introduction

Lung cancer remains the most common cancer worldwide in mortality and the second in incidence [1]. Whereas mortality is decreasing among men, it is increasing among women, partly due to a rising prevalence of female smokers [2, 3]. Epidemiological studies have shown that the clinicopathological features of lung cancer differ between men and women, with a higher percentage of adenocarcinomas, nonsmokers, and EGFR mutations among women [4–7]. In addition, the National Lung Screening Trial [8], the NELSON trial [9], and the LUSI trial [10] identified a layer benefit of screening for lung cancer in women compared to men. The reasons for these differences are not yet clearly understood. Moreover, several studies have found a higher lung cancer risk for smoking women in comparison to smoking men [11, 12]. However, this sex difference does not seem to be related to differences in baseline exposure and smoking history but to the higher susceptibility to tobacco carcinogens in women [5, 13]. The current hypotheses suggest not only the role of estrogens but also the contribution of the genetic makeup [14]. While the first hypothesis has been extensively explored and postulates that estrogens may influence lung-cancer development, either through direct promotion of cell proliferation in the lung or modification of the lung-carcinogen metabolism [14, 15], the genetic hypothesis has not received the same in-depth investigation. Moreover, although many lung cancer susceptibility loci have been identified by genome-wide association studies, these analyses have been restricted to autosomes, without investigating sex differences, and have excluded X-linked loci [16, 17]. Thus, the potential involvement of the sex chromosomes [7, 18] in sex differences in lung cancer incidence, irrespective of smoking habits, remains to be investigated.

The lack of interest for the X chromosome as compared to autosomes is usually explained by a lower coverage of chromosome X, technical issues regarding genotype calling and nonstandard statistical analyses to account for the random/nonrandom X chromosome inactivation process on female X chromosome loci in the non-pseudoautosomal regions [19]. Indeed, the main feature that differentiates the X chromosome from the autosomal chromosomes, apart from the pseudoautosomal regions, is the difference in the number of X chromosome copies between men and women. This dosage imbalance is in part compensated by inactivation of one X chromosome in females. In each female cell, one copy of the X chromosome is inactivated. The X chromosome inactivation (XCI) occurs randomly irrespectively of either the maternal or paternal X chromosome, and very early in embryonic life, and is inherited by all daughter cells through mitosis [20]. Females are a mosaic of 2 cell populations in which either the maternal or paternal alleles of X chromosome genes are expressed. While the random inactivation process results in roughly a symmetrical distribution in most females, nonrandom or skewing of X chromosome inactivation (XCI-S) is observed in some women, leading to a majority of either paternal or maternal XCI [20, 21]. Skewed patterns of XCI probably account for the differential expression of disease phenotypes [22]. In the context of cancer susceptibility, it has been shown that some genetic polymorphisms on the X chromosome have a sex-specific association with cancer risk [23]. Moreover, the skewed XCI has been found to be related to the development of ovarian, breast, and lung cancers [24].

This issue prompted us to perform an association study to search for new lung cancer susceptibility loci on the X chromosome among ever smokers (current and past smokers) from the Canadian population-based cohort CARTaGENE [25]. As our main interest was the occurrence of lung cancer, we performed an association time-to-event analysis where the canonical scale was age. For such time-to-event analysis, we had to take into account nongenetic confounding factors (mainly smoking-related variables) and the potential existence of the nonrandom XCI process among the women.

For time-to-event analyses, the classical approach for investigating the X chromosome considers the well-known Cox proportional hazards model [26], which assumes multiplicative effects of risk factors on the baseline hazard function, and considers either the XCI process or its escape (XCE). To take into account these two underlying biological processes, two genotype coding schemes are commonly used. One corresponds to the assumption of the XCI process as proposed by Clayton [27], while the other corresponds to the XCE process, as implemented in the classical PLINK software [28]. For XCI, the proposed coding values are the same for homozygous females and hemizygous males, while the heterozygous females fall midway between two homozygous, mimicking the fact that about 50% of cells have the minor (or alternative) allele active, while the other 50% of cells have the reference allele active due to random XCI. For XCE, the coding implemented in PLINK codes female genotypes as 0, 1, or 2 copies of the minor allele and male genotypes as 0 or 1 copies of the minor allele. This genotype coding assumes that variants on both copies of the X chromosome are expressed in females. It is only recently that new statistical methods have been proposed to cope with the complex biological process of

nonrandom X-inactivation. By considering XCI as a subject-specific biological process, Xu et al. [29] and Han et al. [30] proposed a penalized partial likelihood approach based on the Cox model with a subject-specific truncated normal latent random variable modeling the XCI-S process.

In this work, we propose to consider an additive-multiplicative hazard model with additive frailty allowing to represent genetic and nongenetic factors in a different way and to account for XCI process. In this framework, the excess risk related to the nongenetic factor is time-related through its multiplicative link to the unknown baseline lung cancer hazard risk, whereas the excess risk due to the genetic factor is constant over time and acts additively. Furthermore, the inclusion of an additive latent (frailty) variable allows to modelize the X-inactivation process. From this framework, we derive a test accounting for random and nonrandom (skewed) XCI and controlling for nongenetic risk factors that play a major role in lung cancer. In the following, we first present the test statistic that is based on a pseudo-score test relying on a semi-parametric additive-multiplicative hazard model with additive frailty. Then, we report the results of a simulation study investigating the properties of the proposed test as compared with the Wald test from a Cox model with random X-inactivation process and the partial likelihood ratio test proposed by Xu et al. accounting for nonrandom X-inactivation process [29, 30]. Then, using the proposed test, we report our findings regarding the association of X chromosome variants with lung cancer susceptibility among smokers in the population-based cohort CARTaGENE, controlling for nongenetic factors (incorporated in the lung cancer risk score PLCOm2012 developed by Tammemägi et al. [31]).

Methods

CARTaGENE Cohort

For this study, we considered the population-based cohort CARTaGENE recruited in phase A (enrolled between 2009 and 2010), composed of 19,985 Quebec residents aged between 40 and 69 years described previously [25]. This cohort consists of men and women residing in metropolitan areas representing a total of 55.7% of the Quebec population (Montreal, Quebec, Sherbrooke, and Saguenay). Participants can be linked with the Quebec administrative health databases, providing data on cancer diagnoses.

Participants Included and Outcome

We included genotyped participants who were smokers or with a history of smoking. Individuals who never smoked or were occasional smokers, who had missing smoking status or a lung cancer diagnosed before recruitment, were excluded.

The outcome of interest was the incidence age of lung cancer. As in the paper by Tonelli et al. [32], we used administrative data to define an incident lung cancer case: individuals with at least two claims in 2 years or one hospitalization related to a lung cancer. The incidence date was the date of first hospital discharge or first claim with the appropriate International Classification of Diseases (9th revision: 1622, 1623, 1624, 1625, 1628, 1629, 2312; 10th revision: C34, D022).

The proposed statistical score test was adjusted for the individual lung cancer risk score proposed by the Tammemägi et al. [31] predictive model (PLCOm2012). This risk score includes the following variables: age, race or ethnic group, education, body mass index, chronic obstructive pulmonary disease (COPD), personal history of cancer, family history of lung cancer, smoking status, smoking intensity, duration of smoking, and smoking quit time. Missing data for variables in the PLCOm2012 model were replaced by the mean values of the PLCOm2012 original article [33] for continuous variables and the mode for categorical variables.

Genetic Data

Around 86% of the participants in phase A had genotyped data. Single-nucleotide polymorphism (SNP) positions were based on build GRCh38. The protocol followed to generate genotyping data is described in online supplementary Material 1 (for all online suppl. material, see https://doi.org/10.1159/000539520).

We extracted 12,236 X chromosome genotyped SNPs from the dataset. The SNPs in the two pseudo-autosomal regions (PAR1, PAR2) were excluded ($n = 389$). As the heterozygote of an SNP in a male should be considered to be a genotyping/calling error, we assigned a missing value for such a call in a male sample. Then we conducted further quality control filtering. The SNPs were removed if they met the following criteria: with genotyping rate <98% for all samples ($n = 712$) or genotyping rate difference >5% between sex ($n = 12$), with minor allele frequency (MAF) <5% for both males and females ($n = 2,651$), deviation from the Hardy-Weinberg equilibrium test with p values $<10^{-5}$ in females ($n = 525$). Finally, a total of 9,261 genotyped individuals who smoked or were past smokers with 8,454 X chromosome SNPs were considered for the subsequent analyses.

Statistical Test

Here, we assume that the nongenetic factors act multiplicatively on the baseline lung cancer incidence, while the genetic information acts additively in a time-independent manner. This model is related to the additive-multiplicative hazard model proposed by Lin and Ying [34] as an extension of their additive model. From this model, we derived a pseudo-score to test the null hypothesis of no effect of the X-linked variants, regardless of the effect of the nongenetic factor.

Notation

To analyze our population-based cohort data, we used a survival model with age as the timescale [35]. Thus, the hazard function can be directly interpreted as the age-specific lung cancer incidence.

Let T denote the time to lung cancer occurrence and C the censoring time. We assume that T and C satisfy the condition of independent and noninformative censoring [36]. For each subject i ($i = 1,\ldots,n$), $X_i = \min(T_i, C_i)$ denotes the observed time of follow-up

and $\delta_i = 1_{(X_i=T_i)}$ the indicator of lung cancer occurrence, where the function $1_{(.)}$ is the indicator function whose value is 1 if the argument is true and 0 otherwise. We also denote $Y_i(t) = 1_{(t \le X_i)}$ the at-risk process and $N_i(t) = 1_{(X_i \le t;\ \delta_i=1)}$ the counting process, given at time *t*.

Let *Z* denote the nongenetic risk factor (here the PLCOm2012 scoring covariate). Let *G* be the genotype for a di-allelic SNP on the X chromosome and denote the two alleles as *A* for the minor (or alternative) and *a* for the major (or reference). Thus, for female subjects, the possible genotypes are *G* = ([aa],[Aa],[AA]) and for male subjects *G* = ([a],[A]).

From the literature, the skewed-inactivation process is frequently arbitrarily defined as a deviation from equal inactivation of each parental allele pattern with the observation of inactivation of the same allele in 75% or more of the cells [21]. In this context, the coded genotypes can be translated to a numeric coding such as *G* = (0,1/2 + *W*,1) for females and *G* = (0,1) for males subjects where *W* is a subject-specific latent (frailty) random variable lying in the interval [−1/2,1/2]. The distribution of this latent variable can be modeled by various continuous probability distributions with support on [−1/2,1/2]. In the following, for each patient *i*, the data consist of $(X_i,\delta_i,G_i,W_i,Z_i)$.

Survival Model

In this work, we use a semi-parametric additive-multiplicative hazard model [34, 37, 38]. The hazard function for the failure time *T* of an individual takes the form:

$$h(t \vee Z = z, G = g, W = w) = h_0(t)e^{\alpha z} + \beta \times \left(g + w1_{(g=1/2)}\right),$$

where $g \in \{0, 1/2, 1\}$, β is the unknown regression coefficient of interest, α is an unknown parameter associated with the scoring PLCOm2012 covariate, and $h_0(t)$ is an unknown and unspecified baseline hazard function. When α is equal to zero, it degenerates to the additive hazard model. When β is equal to zero, it degenerates to the classical Cox model.

Then, the corresponding survival function is such that:

$$S(t \vee Z = z, G = g, W = w) = \exp\left\{-\left\{H_0(t)e^{\alpha z} + \beta\left[g + w \times 1_{(g=1/2)}\right]t\right\}\right\}$$

Here, we assume that for the heterozygous female the *W* are independent and identically distributed random variables [39, 40] lying in the interval [−1/2,1/2] with a probability density function f(w). Based upon this latter assumption, when marginalized over *W*, the unconditional (or marginal) survival function and hazard function are given by:

$$S(t \vee Z = z, G = g) = \exp\left\{-\left\{H_0(t)e^{\alpha z} + \beta g t - 1_{(g=1/2)}\log[L_w(\beta t)]\right\}\right\}$$

$$h(t \vee Z = z, G = g) = h_0(t)e^{\alpha z} + \beta g - 1_{(g=1/2)}\left[\frac{L'_w(\beta t)}{L_w(\beta t)}\right],$$

where $L_w(\beta t)$ and $L'_w(\beta t)$ are the Laplace transform of *W* and its first derivative, respectively.

As the conditional expectation of *W* given $T \ge t$ is such as $E(\beta W \vee T \ge t) = -\left[\frac{L'_w(\beta t)}{L_w(\beta t)}\right]$, we have:

$$h(t \vee Z = z, G = g) = h_0(t)e^{\alpha z} + \beta g + 1_{(g=1/2)}\beta E(W \vee T \ge t)$$

Test

For testing the null hypothesis of no effect of the X-linked variants regardless of the effect of the PLCOm2012 score (H_0:β = 0; ∀α), we considered the same martingale estimating function framework introduced by Lin and Ying [34, 38]. Here, we have the two estimating functions:

$$U_\alpha(t) = \sum_{i=1}^{n}\int_0^\infty (Z_i(t) - \overline{Z}(t))\{dN_i(t) - Y_i(t)\beta G_i dt - Y_i(t)1_{(G_i=\frac{1}{2})}\beta E(W \vee T \ge t)dt\}$$

$$U_\beta(t) = \sum_{i=1}^{n}\int_0^\infty (G_i(t) - \overline{G}(t))\{dN_i(t) - Y_i(t)\beta G_i dt - Y_i(t)\,1_{(G_i=\frac{1}{2})}\beta E\,(W \vee T \ge t)dt\}$$

with $\overline{Z}(t) = \frac{\sum_{i=1}^{n} Y_i(t)e^{\alpha Z_i}Z_i}{\sum_{i=1}^{n} Y_i(t)e^{\alpha Z_i}}$ and $\overline{G}(t) = \frac{\sum_{i=1}^{n} Y_i(t)e^{\alpha Z_i}G_i}{\sum_{i=1}^{n} Y_i(t)e^{\alpha Z_i}}$.

Under the null hypothesis H_0:(β = 0;∀α), we can deduce a pseudo-score statistic where α is replaced by the efficient and consistent maximum partial likelihood estimator $\hat{\alpha}$ computed under the null hypothesis [41].

Then, the final statistic *S* is given by:

$$S = \left(\hat{U}_\beta(t), 0\right)\begin{pmatrix} \hat{A}_{\beta\beta}(t) & \hat{A}_{\beta,\alpha}(t) \\ \hat{A}_{\beta,\alpha}(t) & \hat{A}_{\alpha\alpha}(t) \end{pmatrix}^{-1}\begin{pmatrix} \hat{U}_\beta(t) \\ 0 \end{pmatrix}$$

which is asymptotically distributed under H_0 as a χ^2 with one degree of freedom, with

$$\hat{U}_\beta(t) = \sum_{i=1}^{n}\int_0^\infty (G_i(t) - \overline{G}(t))dN_i(t);$$

$$\hat{A}_{\beta\beta}(t) = \sum_{i=1}^{n}\int_0^\infty (G_i(t) - \overline{G}(t))^2 dN_i(t)$$

$$\hat{A}_{\alpha\alpha}(t) = \sum_{i=1}^{n}\int_0^\infty (Z_i(t) - \overline{Z}(t))^2 dN_i(t);$$

$$\hat{A}_{\beta,\alpha}(t) = \sum_{i=1}^{n}\int_0^\infty (G_i(t) - \overline{G}(t))(Z_i(t) - \overline{Z}(t))dN_i(t)$$

Monte Carlo Simulations

Simulation Procedure

The properties of the proposed test (herein called "pseudo-score") described above were investigated and compared to those of the Wald test from the Cox model using the XCI coding (herein called "Cox-XCI"), the partial likelihood ratio test proposed by Xu et al. based on a random effect XCI-S Cox model (herein called "Xu-LRT").

Data were simulated under various scenarios assuming an additive-multiplicative hazard model: *S*(t∨Z = z, G = g,

Table 1. Size and power of tests (pseudo-score, Cox-XCI, Xu-LRT) for truncated normal distributions (threshold level of 0.05)

Truncated normal distribution with E(X) = 0; cens = 0%			
p_A = 0.10 β:	0	−1	−2
Pseudo-score	0.06	0.33	0.91
Cox-XCI	0.05	0.27	0.88
Xu-LRT	0.1	0.35	0.91
p_A = 0.30 β:	0	−1	−2
Pseudo-score	0.04	0.34	0.91
Cox-XCI	0.04	0.28	0.85
Xu-LRT	0.08	0.36	0.91
Truncated normal distribution with E(X) = 0; cens = 40%			
p_A = 0.10 β:	0	−1	−2
Pseudo-score	0.05	0.23	0.68
Cox-XCI	0.04	0.18	0.59
Xu-LRT	0.09	0.24	0.66
p_A = 0.30 β:	0	−1	−2
Pseudo-score	0.05	0.3	0.88
Cox-XCI	0.05	0.26	0.84
Xu-LRT	0.09	0.33	0.89
Truncated normal distribution with E(X) = −0.25; cens = 0%			
p_A = 0.10 β:	0	−1	−2
Pseudo-score	0.06	0.3	0.85
Cox-XCI	0.05	0.25	0.81
Xu-LRT	0.1	0.32	0.86
p_A = 0.30 β:	0	−1	−2
Pseudo-score	0.06	0.32	0.83
Cox-XCI	0.06	0.26	0.79
Xu-LRT	0.11	0.35	0.83
Truncated normal distribution with E(X) = −0.25; cens = 40%			
p_A = 0.10 β:	0	−1	−2
Pseudo-score	0.05	0.23	0.6
Cox-XCI	0.06	0.16	0.51
Xu-LRT	0.1	0.23	0.59
p_A = 0.30 β:	0	−1	−2
Pseudo-score	0.05	0.24	0.83
Cox-XCI	0.05	0.21	0.79
Xu-LRT	0.09	0.29	0.84

Table 2. Size and power of tests (pseudo-score, Cox-XCI, Xu-LRT) for Beta distributions (threshold level of 0.05)

Shifted Beta distribution with E(X) = 0; cens = 0%			
p_A = 0.10 β:	0	−1	−2
Pseudo-score	0.05	0.3	0.9
Cox-XCI	0.06	0.26	0.87
Xu-LRT	0.09	0.32	0.9
p_A = 0.30 β:	0	−1	−2
Pseudo-score	0.05	0.33	0.88
Cox-XCI	0.06	0.26	0.87
Xu-LRT	0.1	0.36	0.89
Shifted Beta distribution with E(X) = 0; cens = 40%			
p_A = 0.10 β:	0	−1	−2
Pseudo-score	0.05	0.2	0.69
Cox-XCI	0.04	0.16	0.59
Xu-LRT	0.09	0.22	0.67
p_A = 0.30 β:	0	−1	−2
Pseudo-score	0.05	0.28	0.88
Cox-XCI	0.06	0.26	0.85
Xu-LRT	0.1	0.33	0.9
Shifted Beta distribution with E(X) = −0.25; cens = 0%			
p_A = 0.10 β:	0	−1	−2
Pseudo-score	0.06	0.31	0.89
Cox-XCI	0.06	0.25	0.87
Xu-LRT	0.1	0.33	0.9
p_A = 0.30 β:	0	−1	−2
Pseudo-score	0.05	0.28	0.89
Cox-XCI	0.05	0.22	0.87
Xu-LRT	0.1	0.31	0.9
Shifted Beta distribution with E(X) = −0.25; cens = 40%			
p_A = 0.10 β:	0	−1	−2
Pseudo-score	0.05	0.18	0.68
Cox-XCI	0.04	0.14	0.6
Xu-LRT	0.08	0.19	0.66
p_A = 0.30 β:	0	−1	−2
Pseudo-score	0.05	0.26	0.89
Cox-XCI	0.05	0.22	0.85
Xu-LRT	0.09	0.3	0.88

W = w) = exp[−(λ_0(t)$e^{\alpha z}$+ β × (g + w)t)] with *G* a locus undergoing XCI, *Z* a continuous confounding covariate, and *W* a latent variable. In practice, genotype information for females was generated by combining the values of two Bernoulli variables (B($p_{[A]}$)) independently drawn and for males from only one Bernoulli variable with mean: 15% and 30% for both males and females. This value corresponds to a pseudo-MAF, i.e., the proportion of [*A*] allele in the simulated population. The ratio between the female and male rate was set to 1:1. The confounding variable *Z* was generated from a standard normal distribution. The parameter α was set to 0.2 and $\lambda_0(t)$ to 5. The parameter values for β were set to: 0, −1, −2 (protective effect of the minor allele).

Four distributions for the latent variable *W* were investigated and generated independently and identically from: (i) a truncated normal distribution ranging from −0.5 to 0.5 with mean zero and standard error of 0.2; (ii) a truncated normal distribution ranging from −0.5 to 0.5 with mean −0.25 and standard error of 0.2; (iii) a Beta distribution with mean 1/2 with a shift value of −0.5; and (iv) a Beta distribution with mean 0.25 with a shift value of −0.5.

We investigated no censoring and 40% type I censoring (administrative censoring). The total number of subjects was chosen to be 400. For each configuration of parameters, 1,000 replications were performed and the levels and powers of the tests were estimated with a 0.05 significance level.

Results

Simulation Study

Tables 1 and 2 display the results of the simulations under the four scenarios (Beta and truncated normal distributions). The estimated levels of the proposed test and the Cox-XCI test under the null hypothesis are within the binomial range [0.0365–0.0635] in each configuration. This is not the case for the Xu-LRT test, which shows a substantial type I error inflation in each configuration.

Online supplementary Figures S1, S2, and S3 display the histograms of the *p* values under the null hypothesis (β = 0), no censoring and Beta distribution for the latent variable together with the QQ plot of the *p* values (on the –log10 scale with the confidence bands). As seen from these figures, the distribution of the *p* values for the Xu-LRT test shows a marked departure from uniform distribution. This is not the case for the proposed test and the Cox-XCI test.

The power of the proposed test and the Xu-LRT test is very close. However, as compared to the proposed test, the Xu et al. test showed inflated type I error rate for all the scenarios. The Cox is always less powerful than our proposed test. As expected, for each test, the performance decreases with the percentage of censoring and increases with the MAF. The power results seem not to be sensibly modified by the distribution of the random effect.

X-WAS in CARTaGENE Cohort

The cohort analyzed in this work consists in 9,261 individuals with genotyping information who successfully passed the quality-control procedure and who smoked or were past smokers. The baseline characteristics are presented in Table 3. There were 4,529 women (48.9%) and 4,732 men (51.1%). Patients were aged from 40.2 to 70.3 years old and the median age was 54.2 years old. The men were slightly older (55.0 vs. 53.5, $p < 0.001$), had a higher body mass index ($p < 0.001$), had a higher income ($p < 0.001$), and had a higher educational level (graduate/university 19.9 vs. 16.5%, $p < 0.001$). The women declared more cancer history (10.7 vs. 7.4%, $p < 0.001$) and COPD history (9.2 vs. 6.1%, $p < 0.001$). Among the women, 1,601 (35.8%) had reached menopause and were not treated by hormone replacement therapy.

The median follow-up after cohort entry was 5.83 years. A higher PLCOm2012 score was observed in men (median: 0.50 vs. 0.44%; $p = 0.001$). When considering the classical threshold of 1.51% for screening, 16.5% of individuals were at a higher risk of lung cancer (PLCOm2012 score ≥1.51%). The proportion was significantly higher in men than in women (17.6 vs. 15.4%, $p = 0.004$). Figure 1 shows the risk score distribution of the PLCOm2012 model. In this cohort, 150 (1.62%) participants suffered lung cancer after inclusion (1.73% males, 1.50% females, $p = 0.42$). Using a univariate Cox model (Table 4), the PLCOm2012 risk score was significantly associated with the occurrence of lung cancer (HR 1.16 [1.11; 1.21]; $p < 0.001$) but not sex ($p = 0.66$). COPD history, cancer history, and the four smoking exposure variables were associated with the occurrence of lung cancer ($p < 0.001$). Among women, the menopause and the use of hormone replacement therapy were not associated with the incidence of a lung cancer ($p = 0.84$ and $p = 0.33$, respectively).

We reported the results obtained with the proposed test together with those obtained from the classical Wald test from the Cox model (under XCI process) and the Xu et al.'s XCI and XCI-S likelihood ratio tests [29, 30]. To select X-chromosomal loci associated with lung cancer and accounting for multiple comparisons, we performed a family-wise error rate-based X-genome-wide analysis with Bonferroni correction. Controlling the family-wise error rate at the nominal level of 5% for 8,454 SNPs, we selected two associated signals: rs12558491 and rs12835699. The first

Table 3. Cohort baseline characteristics

	All (*N* = 9,261)	Women (*N* = 4,529)	Men (*N* = 4,732)	*p* value
Age	54.2 [49.0; 61.3]	53.5 [48.8; 60.1]	55.0 [49.3; 62.3]	<0.001
City				0.919
Laval, *n* (%)	752 (8.12)	362 (7.99)	390 (8.24)	
Montreal, *n* (%)	3,678 (39.7)	1,794 (39.6)	1,884 (39.8)	
North Shore (Montreal), *n* (%)	1,269 (13.7)	624 (13.8)	645 (13.6)	
Quebec, *n* (%)	1,368 (14.8)	661 (14.6)	707 (14.9)	
Saguenay, *n* (%)	337 (3.64)	176 (3.89)	161 (3.40)	
Sherbrooke, *n* (%)	473 (5.11)	230 (5.08)	243 (5.14)	
South Shore (Montreal), *n* (%)	1,384 (14.9)	682 (15.1)	702 (14.8)	
Household income per year				<0.001
<50,000 $, *n* (%)	3,402 (39.1)	1,758 (41.8)	1,644 (36.5)	
50,000–99,999 $, *n* (%)	3,195 (36.7)	1,499 (35.7)	1,696 (37.7)	
>100,000 $, *n* (%)	2,103 (24.2)	944 (22.5)	1,159 (25.8)	
Lung cancer incidence, *n* (%)	150 (1.62)	68 (1.50)	82 (1.73)	0.424
Follow-up time, years	60.0 [54.8; 67.1]	59.3 [54.6; 65.9]	60.8 [55.1; 68.1]	<0.001
PLCOm2012 absolute risk, %	0.47 [0.20; 1.04]	0.44 [0.17; 0.99]	0.50 [0.22; 1.08]	<0.001
PLCOm2012 absolute risk higher than 1.51%, *n* (%)	1,531 (16.5)	697 (15.4)	834 (17.6)	0.004
Menopause, *n* (%)	2,811 (62.8)	2,811 (62.8)	–	–
Menopause without hormone replacement therapy, *n* (%)	1,601 (35.8)	1,601 (35.8)	–	–
Hormone replacement therapy, *n* (%)	1,334 (29.6)	1,334 (29.6)	–	–
Highest level of education				<0.001
Less than high-school graduate, *n* (%)	245 (2.65)	106 (2.35)	139 (2.94)	
High-school graduate, *n* (%)	2,659 (28.8)	1,331 (29.5)	1,328 (28.1)	
Some college, *n* (%)	3,025 (32.7)	1,550 (34.3)	1,475 (31.2)	
College graduate, *n* (%)	1624 (17.6)	785 (17.4)	839 (17.8)	
Postgraduate or professional degree, *n* (%)	1,686 (18.2)	746 (16.5)	940 (19.9)	
Body mass index	27.0 [24.2; 30.4]	26.1 [23.1; 30.0]	27.7 [25.2; 30.7]	<0.001
COPD history, *n* (%)	701 (7.61)	416 (9.22)	285 (6.07)	<0.001
Cancer history, *n* (%)	831 (9.00)	481 (10.7)	350 (7.42)	<0.001
Family history of lung cancer, *n* (%)	1,236 (13.7)	623 (14.1)	613 (13.4)	0.385
Smoking status				0.323
Daily smoker, *n* (%)	2,428 (26.2)	1,166 (25.7)	1,262 (26.7)	
Past smoker, *n* (%)	6,833 (73.8)	3,363 (74.3)	3,470 (73.3)	
Current smokers, cigarettes/day	17.0 [8.00; 23.0]	13.0 [8.00; 20.0]	18.0 [13.0; 23.0]	<0.001
Past smokers, cigarettes/day	18.0 [13.0; 23.0]	18.0 [8.00; 23.0]	18.0 [13.0; 23.0]	<0.001
Smoking duration, years	23.0 [12.0; 33.0]	23.0 [11.0; 32.9]	23.4 [13.0; 33.2]	<0.001
Smoking quit duration, years	19.9 [10.1; 27.9]	19.6 [9.83; 27.0]	20.2 [10.4; 28.7]	0.001

SNP rs12558491 was significantly associated with lung cancer ($p = 2.75\times10^{-9}$) and is located in the cytoband Xp21.3-p21.2 in *IL1RAPL1* (IL-1 receptor accessory protein-like 1) gene. The second SNP rs12835699, located in the same gene, was also significantly associated with lung cancer ($p = 1.26\times10^{-6}$). These two intronic SNPs were in low linkage disequilibrium ($r^2 = 0.25$) [42]. No signal was detected using the classical Cox multiplicative hazards model adjusted for the PLCOm2012 risk score under the random XCI process. In addition, no signal was detected using the Xu et al.'s likelihood ratio test under XCI and XCI-S [29, 30].

For the SNP rs12558491, the frequency of the minor allele was 5.93% for females and 5.83% for males, with no significant sex difference. For males, there were 276 hemizygous for the minor allele and 4,456 for the major allele. For females, there were 14 homozygous for the minor allele, 4,006 for the major allele, and 509 heterozygous.

For the SNP rs12835699, the frequency for the minor allele was 5.83% for females and 5.95% for males with no significant difference between sexes. For the males, there were 275 hemizygous for the minor allele and 4,348 for the major allele. For the females, there were 16 homozygous for the minor allele, 4,005 for the major allele, and 495 heterozygous.

Online supplementary Figure S4 shows the Manhattan plot of the X chromosome genome-wide association results obtained with the proposed test statistic, the Wald test under the Cox model (under XCI process), and the XCI-S

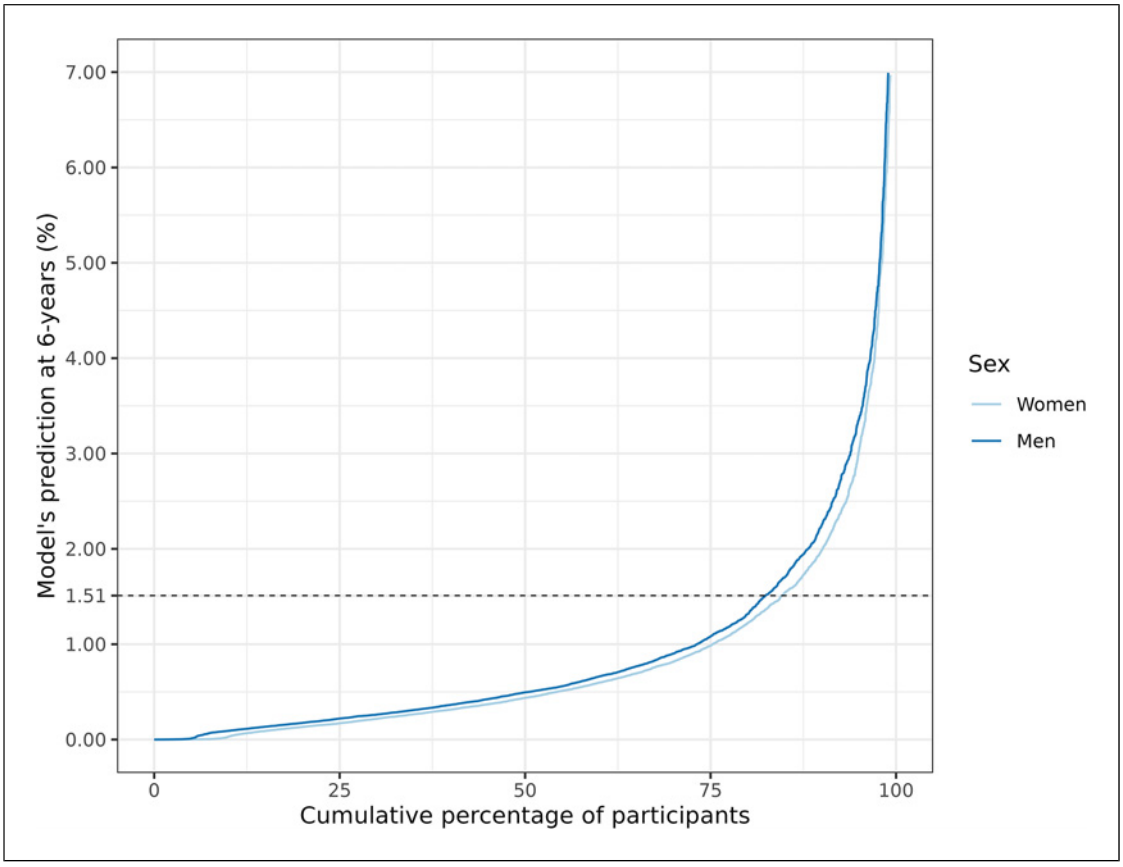

Fig. 1. Absolute risk score distribution of PLCOm2012 model for men and women.

likelihood ratio tests [29, 30]. Figure 2 shows the Kaplan-Meier survival curves for the SNP rs12558491. No event occurred among the 276 hemizygous males and 14 homozygous females for the minor allele. Among the hemizygous males and homozygous females for the major allele, 82 and 64 individuals suffered lung cancer, respectively. Among the heterozygous females, 4 individuals suffered lung cancer. Figure 3 displays the Kaplan-Meier survival curves grouped along the genotype of the SNP rs12558491 and the high-/low-risk score status with the 1.51% threshold.

Online supplementary Figure S5 shows the Kaplan-Meier survival curves for the SNP rs12835699. No event occurred among the 275 hemizygous males and 16 homozygous females for the minor allele. Among the hemizygous males and homozygous females for the major allele, 81 and 63 individuals suffered lung cancer, respectively. Among the heterozygous females, 5 individuals suffered lung cancer. Online supplementary Figure S6 shows the Kaplan-Meier survival curves grouped along the genotype of the SNP rs12835699 and the high-/low-risk score status with the 1.51% threshold.

Discussion

Investigating sex differences in the risk of developing cancer is an important issue that can help to unravel the processes driving carcinogenesis and to define subgroups at the highest risk. In this context, X chromosome-wide association studies (X-WAS) are of considerable interest but are seldom performed. Thus, we performed an X-WAS for lung cancer using a new statistical testing strategy accounting for random/nonrandom (skewed) XCI process while adjusting for nongenetic factors. To do so, we proposed a pseudo-score test relying on a semi-parametric additive-multiplicative hazard model. For time-to-event analysis, the multiplicative Cox proportional and additive hazards models represent the two main frameworks for studying the association between risk factor and time-to-event [37]. However, they both suppose that the true underlying covariate effects are either additive or multiplicative on the baseline hazards, but not both. Here, we considered an additive-multiplicative hazards model that is simple, easily interpretable and allows genetic

Table 4. Univariate Cox model results

	No lung cancer (*N* = 9,111)	Lung cancer (*N* = 150)	Hazard ratio	*p* value
Sex				0.658
Women	4,461 (49.0%)	68 (45.3%)	Ref.	
Men	4,650 (51.0%)	82 (54.7%)	0.93 [0.67; 1.28]	
PLCOm2012 absolute risk, %	0.46 [0.19; 1.01]	1.42 [0.40; 3.81]	1.16 [1.11; 1.21]	<0.001
PLCOm2012 absolute risk higher than 1.51%	1,457 (16.0%)	74 (49.3%)	2.15 [1.55; 2.98]	<0.001
Household income per year				0.085
<50,000 $	3,323 (38.8%)	79 (56.0%)	Ref.	
50,000–USD 99,999 $	3,155 (36.9%)	40 (28.4%)	0.66 [0.45; 0.97]	
>100,000 $	2,081 (24.3%)	22 (15.6%)	0.75 [0.46; 1.20]	
Highest level of education				0.536
Less than high-school graduate	237 (2.61%)	8 (5.33%)	Ref.	
High-school graduate	2,608 (28.7%)	51 (34.0%)	1.02 [0.48; 2.15]	
Some college	2,987 (32.9%)	38 (25.3%)	0.75 [0.35; 1.61]	
College graduate	1,597 (17.6%)	27 (18.0%)	0.94 [0.43; 2.07]	
Postgraduate or professional degree	1,660 (18.3%)	26 (17.3%)	0.74 [0.34; 1.64]	
Body mass index	27.0 [24.2; 30.4]	26.9 [23.2; 31.5]	1.00 [0.97; 1.03]	0.925
COPD history	666 (7.35%)	35 (23.6%)	3.31 [2.27; 4.84]	<0.001
Cancer history	797 (8.78%)	34 (23.0%)	Ref.	<0.001
Family history of lung cancer	1,210 (13.7%)	26 (18.2%)	1.29 [0.84; 1.97]	0.245
Smoking status				<0.001
Daily smoker	2,368 (26.0%)	60 (40.0%)	Ref.	
Past smoker	6,743 (74.0%)	90 (60.0%)	0.30 [0.22;0.42]	
Current smokers, cigarettes/day	15.0 [8.00; 23.0]	18.0 [15.0; 24.5]	1.04 [1.02; 1.06]	<0.001
Past smokers, cigarettes/day	18.0 [13.0; 23.0]	23.0 [18.0; 38.0]	1.03 [1.02; 1.04]	<0.001
Smoking duration, years	23.0 [12.0; 33.0]	35.0 [21.0; 44.3]	1.03 [1.02; 1.05]	<0.001
Smoking quit duration, years	19.9 [10.1; 28.0]	19.3 [10.3; 24.2]	0.96 [0.95; 0.98]	<0.001

and nongenetic factors to be considered in a different way [34, 38]. With this model, the nongenetic factor is time-related through its multiplicative link to the unknown baseline lung cancer hazard risk, whereas the genetic factor is constant over time. Moreover, an interesting feature of this approach is that when the hazard is marginalized over the latent variable (related to the XCI process), the genetic part is still additive. Based on this model, a simple test for the absence of a genetic effect can be deduced and easily implemented with standard softwares. In our context, the interest of this modelization is that it provides a way of including modifiable time-related factors (mainly tobacco-related) together with nonmodifiable time-unrelated factor (genetic). Obviously, this model may have difficulties representing data with time-varying effect for the additive part. Other models that combine multiplicative and additive part could be considered in further works.

The simulation study shows that the power gains of the proposed test and the Xu et al. test are very close, but the inflated type I error rate of the Xu et al. test does not argue in favor of this latter test since the best one is the one providing maximal power while maintaining a type I error rate at the nominal level. We investigated various scenarios that led to similar results. We may hypothesize that the good behavior of the proposed test as compared to the Xu et al.'s XCI-S test is probably related to the fact that the proposed test is obtained as a pseudo-score test without any estimation of the random effect, whereas the Xu et al.'s XCI-S test is a likelihood ratio test that requires the estimation of the random effect from the penalized partial likelihood.

As our results demonstrate, this test seems effective for analyzing X-WAS, since the interesting signals on the X chromosome would have been overlooked by examining

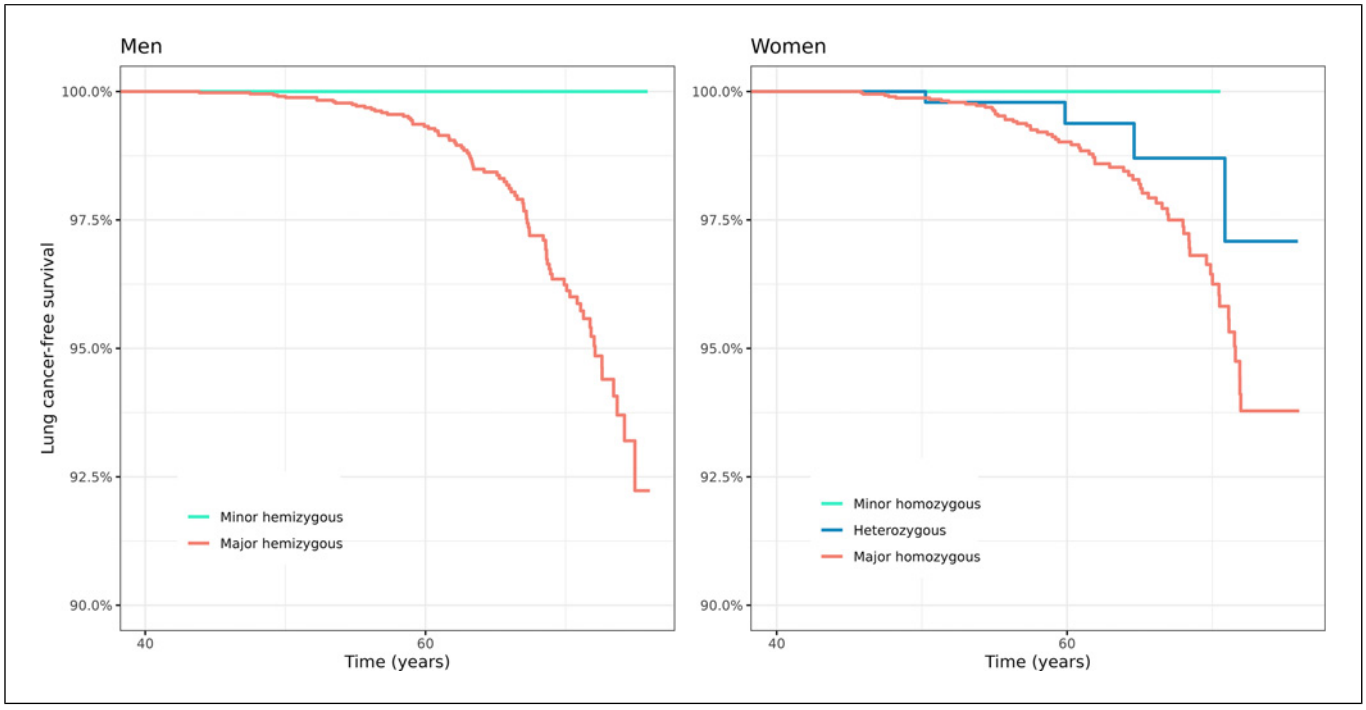

Fig. 2. Kaplan-Meier survival curves for men and women (SNP rs12558491).

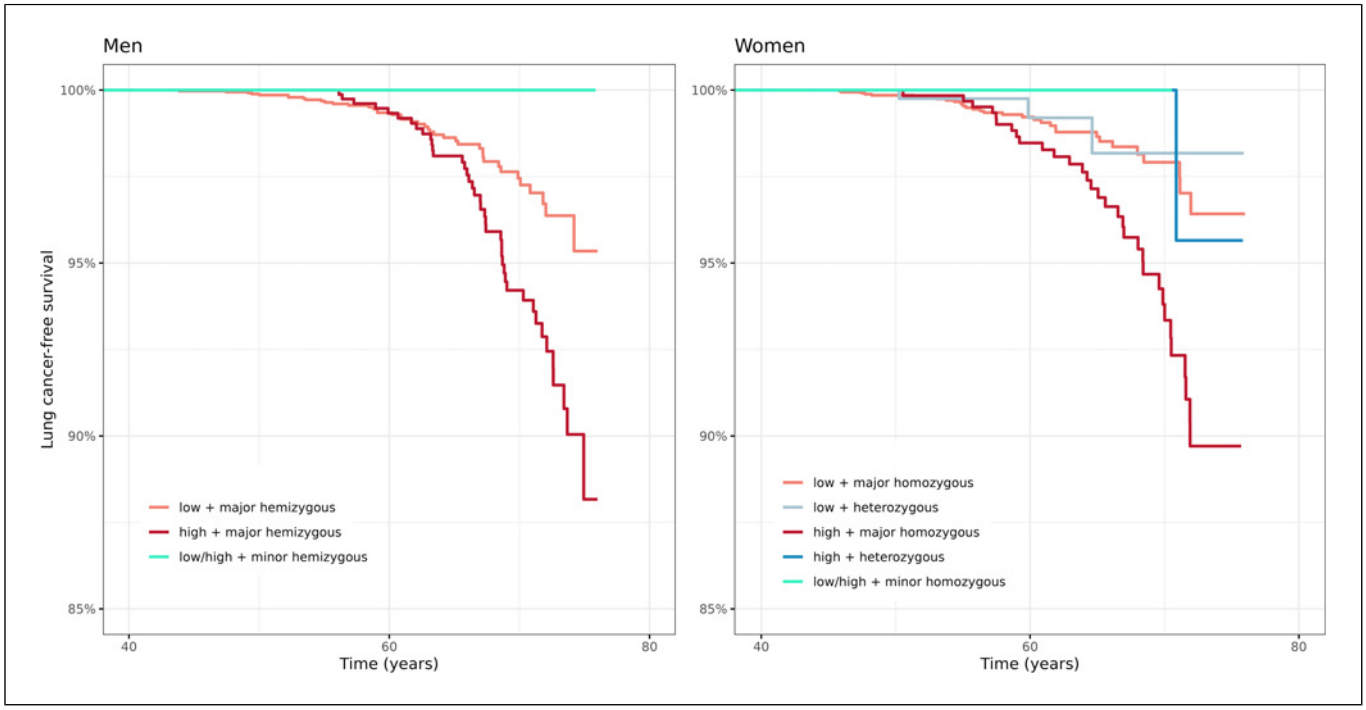

Fig. 3. Kaplan-Meier survival curves for men and women grouped along genotype (SNP rs12558491) and low-/high-risk score status (PLCOm2012 risk 1.5% threshold).

only results from the Cox model under the XCI process or the Xu et al. XCI/XCI-S likelihood ratio test derived under a multiplicative hazard survival model [29, 30]. The difference between these latter tests and the proposed test is that the genetic factor in our model is assumed to be constant over time and additive, which seems to reflect the actual relationship more accurately than a multiplicative model.

We analyzed the data from the population-based Canadian cohort CARTaGENE composed of middle-aged and older adults. In total, we had 9,261 smokers or past smokers. As expected, the PLCOm2012 risk score, which was higher for men than for women, was strongly associated with the occurrence of a lung cancer. Among its components, COPD history, cancer history, and the four smoking exposure variables were significantly associated with the occurrence of a lung cancer. Of note, women declared more cancer and COPD history. No relationship with lung cancer was found with either sex or hormonal status.

Adjusting for the nongenetic risk score and taking into account for multiple testing, two SNPs (rs12558491, rs12835699) were selected. These two signals, in low linkage disequilibrium, were observed in the IL1RAPL1 gene, suggesting its putative role. The minor allele for these two SNPs was associated with a lower incidence of lung cancer in the homozygous females and hemizygous males. The IL1RAPL1 gene, which is located in a fragile site, is highly expressed in the brain, and some gene alterations (deletions and mutations) have been reported in patients with intellectual disability. Moreover, non-random inactivation of large common fragile site genes has also been described in other diseases such as cancers [43]. IL1RAPL1 is an orphan receptor and a member of the interleukin-1 (IL-1) receptor family. The IL-1 family of cytokines and their receptors are a heterogeneous group of proteins that regulate immunity. While its ligand is not known and its downstream signaling is poorly understood, IL1RAPL1 has been shown to mediate the anti-inflammatory function of the cytokine IL-38. IL1-RAPL1 is one of the three receptors (with IL-1R1 and IL-36R), associated with the IL-38 pathway, one of the more recently reported and least understood cytokines of the IL-1 family [44].

The exact biological function of IL-38 remains controversial and seems context-dependent. IL-38 has an anti-inflammatory role [45] and is mainly involved in innate immunity. After tissue damage, IL-38 is released by various cells including epithelial cells, immune cells (including B cells, dendritic cells, and macrophages), and endothelial cells to decrease inflammation. The anti-inflammatory role of IL-38 has been demonstrated in various human inflammatory and autoimmune diseases, and polymorphisms of the IL-38 gene have been reported to be associated with autoimmune arthritis [45]. Furthermore, in a recent case-control X-WAS among Latin American children, a variant of the IL1RAPL1 gene was found to be negatively associated with asthma [46]. The anti-inflammatory/tissue protective function of IL-38 is mediated by binding to a cell surface receptor of the IL-1 receptor family, including IL1RAPL1. There is evidence that IL1RAPL1 is involved in innate immune cell activation [47] through the activation of the JNK/AP1 pathway. It has been shown that a mature, truncated form of IL-38 is produced by apoptotic cells in order to limit inflammation in injured tissues. Truncated IL-38 binds to IL1RAPL1 and reduced IL-6 production by attenuating the JNK/AP1 pathway downstream of IL1RAPL1 [45]. This mechanism may be relevant in the resolution of inflammation, autoimmunity, and cancer.

In the lung parenchyma, the anti-inflammatory role of IL-38 has been demonstrated in both in vitro and in vivo models [44]. IL-38 is barely detectable in the normal lung, but is overexpressed in various chronic inflammatory conditions such as interstitial lung diseases. Its expression is mainly evident in the hyperplastic type II pneumocytes, i.e., hallmark of regenerative epithelial changes following alveolar damage, in keeping with a tissue-protective effect [47]. Cigarette smoke induces chronic inflammation and epithelial cell damage, resulting in COPD, emphysema, and structural alterations such as an increase in cell proliferation, angiogenesis, and apoptosis arrest that may lead to tumor growth, raising the question of the role of IL-38/IL1RAPL1 pathway in lung carcinogenesis. Moreover, IL-38 seems to play a pro-tumoral role in lung cancer, with a high expression associated with high-grade tumors, poor survival, and PD-L1 expression. It is thought that IL-38 could shape an immunosuppressive tumor microenvironment favorable to tumor growth [44]. The exact role of IL1RAPL1 in lung cancer is not yet known, but it is the receptor with the largest body of evidence linking it to IL-38 signaling [44].

To our knowledge, this is the first cohort study investigating the role of X chromosome in lung cancer susceptibility. One of its strengths is that the investigation was performed on a well-defined population-based cohort and that we used a test statistic that takes XCI-S into account and adjusts for nongenetic risk factors. Nevertheless, it also has some limitations. First, our cohort is an open cohort, in which a subject's lung cancer occurrence is conditional to the fact that he/she was

free of disease at his/her age of recruitment between 39 and 69 years old. Even though this window of time is limited, it represents a period of time where the incidence rate of lung cancer is high. Moreover, since we used age as the timescale rather than time since recruitment (time-on-study), the hazard function can be directly interpreted as the age-specific incidence function for lung cancer. Second, some variables on the PLCOm2012 were self-reported and we cannot rule out sex differences in self-reporting. Third, we tested for a lack of association rather than focusing on estimation, so the main drawback is that the proposed statistic might not be fully efficient when the underlying model is incorrect. However, the other tests seem less appropriate and experienced more power loss. Fourth, even though the random/nonrandom XCI process silences most genes on one X chromosome in females, some genes may escape XCI and may have been missed by our analysis.

In this study, we performed an X chromosome-wide association study on lung cancer, taking into account smoking behavior and X-inactivation process, using a Canadian population-based cohort. We identified two loci associated with lung cancer located in the IL1RAPL1 gene. For both SNPs, the minor allele is associated with a lower risk of lung cancer. Further investigations are needed to validate this association. This study underlines the need for more X-WAS studies using well-designed statistical tests.

Acknowledgments

We would like to thank all the CARTaGENE participants for their generous investments in health research. We would also like to thank the RAMQ and the Commission d'accès à l'information (CAI) for their support in obtaining the data.

Statement of Ethics

This study protocol was reviewed and approved by the Research Ethics Board of the CHU Sainte-Justine (Approval No. 2020-2427). In addition, CARTaGENE obtained ethics approval by the CHU Sainte-Justine under the reference: MP-21-2011-345, 3297. Written informed consent was obtained for participation in this study (cartagene.qc.ca/files/documents/consent/brochure_en_0505.pdf).

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Funding Sources

The "*CHU Sainte-Justine Research Center*" funded Rodolphe Jantzen. The "*Fonds de Recherche du Québec – Santé*" funded Nicole Ezer. The funder had no role in the design, data collection, data analysis, and reporting of this study.

Author Contributions

R.J.: conceptualization, data curation, formal analysis, investigation, visualization, and writing. N.E. and S.C.B.: writing. P.B: conceptualization, formal analysis, methodology, project administration, supervision, validation, and writing. All authors read and approved the final manuscript.

Data Availability Statement

The data that support the findings of this study are available from CARTaGENE, but restrictions apply to their availability. Data are, however, available directly from CARTaGENE (http://cartagene.qc.ca; access@cartagene.qc.ca; +1 514-345-2156). Further inquiries can be directed to the corresponding author.

References

1 Barta JA, Powell CA, Wisnivesky JP. Global epidemiology of lung cancer. Ann Glob Health. 2019;85(1):8. https://doi.org/10.5334/aogh.2419

2 Fidler-Benaoudia MM, Torre LA, Bray F, Ferlay J, Jemal A. Lung cancer incidence in young women *vs.* young men: a systematic analysis in 40 countries. Int J Cancer. 2020;147(3):811–9. https://doi.org/10.1002/ijc.32809

3 Jemal A, Miller KD, Ma J, Siegel RL, Fedewa SA, Islami F, et al. Higher lung cancer incidence in young women than young men in the United States. N Engl J Med. 2018;378(21):1999–2009. https://doi.org/10.1056/nejmoa1715907

4 Yin X, Zhu Z, Hosgood HD, Lan Q, Seow WJ. Reproductive factors and lung cancer risk: a comprehensive systematic review and meta-analysis. BMC Publ Health. 2020;20(1):1458. https://doi.org/10.1186/s12889-020-09530-7

5 Mederos N, Friedlaender A, Peters S, Addeo A. Gender-specific aspects of epidemiology, molecular genetics and outcome: lung cancer. ESMO Open. 2020;5:e000796. https://doi.org/10.1136/esmoopen-2020-000796

6 Zhu Y, Shao X, Wang X, Liu L, Liang H. Sex disparities in cancer. Cancer Lett. 2019;466:35–8. https://doi.org/10.1016/j.canlet.2019.08.017

7 Shi X, Young S, Morahan G. Identification of genetic variants associated with sex-specific lung-cancer risk. Cancers. 2021;13(24):6379. https://doi.org/10.3390/cancers13246379

8 The National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365(5):395–409. https://doi.org/10.1056/nejmoa1102873

9 de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA, et al. Reduced lung-cancer mortality with volume CT screening in a randomized trial. N Engl J Med. 2020;382(6):503–13. https://doi.org/10.1056/nejmoa1911793

10 Becker N, Motsch E, Trotter A, Heussel CP, Dienemann H, Schnabel PA, et al. Lung cancer mortality reduction by LDCT screening: results from the randomized German LUSI trial. Int J Cancer. 2020;146(6):1503–13. https://doi.org/10.1002/ijc.32486
11 Papadopoulos A, Guida F, Leffondré K, Cénée S, Cyr D, Schmaus A, et al. Heavy smoking and lung cancer: are women at higher risk? Result of the ICARE study. Br J Cancer. 2014;110(5):1385–91. https://doi.org/10.1038/bjc.2013.821
12 International Early Lung Cancer Action Program Investigators. Women's susceptibility to tobacco carcinogens and survival after diagnosis of lung cancer. JAMA. 2006;296(2):180–4. https://doi.org/10.1001/jama.296.2.180
13 Ben-Zaken Cohen S, Paré PD, Man SFP, Sin DD. The growing burden of chronic obstructive pulmonary disease and lung cancer in women. Am J Respir Crit Care Med. 2007;176(2):113–20. https://doi.org/10.1164/rccm.200611-1655pp
14 Siegfried JM. Women and lung cancer: does oestrogen play a role? Lancet Oncol. 2001;2(8):506–13. https://doi.org/10.1016/s1470-2045(01)00457-0
15 Stapelfeld C, Dammann C, Maser E. Sex-specificity in lung cancer risk. Int J Cancer. 2020;146(9):2376–82. https://doi.org/10.1002/ijc.32716
16 Lopes-Ramos CM, Quackenbush J, DeMeo DL. Genome-wide sex and gender differences in cancer. Front Oncol. 2020;10. https://doi.org/10.3389/fonc.2020.597788
17 Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Res. 2023;51(D1):D977–85. https://doi.org/10.1093/nar/gkac1010
18 Ragavan M, Patel MI. The evolving landscape of sex-based differences in lung cancer: a distinct disease in women. Eur Respir Rev. 2022;31(163):210100. https://doi.org/10.1183/16000617.0100-2021
19 Wise AL, Gyi L, Manolio TA. eXclusion: toward integrating the X chromosome in genome-wide association analyses. Am J Hum Genet. 2013;92(5):643–7. https://doi.org/10.1016/j.ajhg.2013.03.017
20 Shvetsova E, Sofronova A, Monajemi R, Gagalova K, Draisma HHM, White SJ, et al. Skewed X-inactivation is common in the general female population. Eur J Hum Genet. 2019;27(3):455–65. https://doi.org/10.1038/s41431-018-0291-3
21 Minks J, Robinson WP, Brown CJ. A skewed view of X chromosome inactivation. J Clin Invest. 2008;118(1):20–3. https://doi.org/10.1172/jci34470
22 Ober C, Loisel DA, Gilad Y. Sex-specific genetic architecture of human disease. Nat Rev Genet. 2008;9(12):911–22. https://doi.org/10.1038/nrg2415
23 Achilla C, Papavramidis T, Angelis L, Chatzikyriakidou A. The implication of X-linked genetic polymorphisms in susceptibility and sexual dimorphism of cancer. Anticancer Res. 2022;42(5):2261–76. https://doi.org/10.21873/anticanres.15706
24 Li G, Su Q, Liu G-Q, Gong L, Zhang W, Zhu S-J, et al. Skewed X chromosome inactivation of blood cells is associated with early development of lung cancer in females. Oncol Rep. 2006;16(4):859–64. https://doi.org/10.3892/or.16.4.859
25 Awadalla P, Boileau C, Payette Y, Idaghdour Y, Goulet J-P, Knoppers B, et al. Cohort profile of the CARTaGENE study: Quebec's population-based biobank for public health and personalized genomics. Int J Epidemiol. 2013;42(5):1285–99. https://doi.org/10.1093/ije/dys160
26 Cox DR. Regression models and life-tables. J R Stat Soc Ser B Methodol. 1972;34(2):187–202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x
27 Clayton D. Testing for association on the X chromosome. Biostatistics. 2008;9(4):593–600. https://doi.org/10.1093/biostatistics/kxn007
28 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–75. https://doi.org/10.1086/519795
29 Xu W, Hao M. A unified partial likelihood approach for X-chromosome association on time-to-event outcomes. Genet Epidemiol. 2018;42(1):80–94. https://doi.org/10.1002/gepi.22097
30 Han D, Hao M, Qu L, Xu W. A novel model for the X-chromosome inactivation association on survival data. Stat Methods Med Res. 2020;29(5):1305–14. https://doi.org/10.1177/0962280219859037
31 Tammemägi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, et al. Selection criteria for lung-cancer screening. N Engl J Med. 2013;368(8):728–36. https://doi.org/10.1056/nejmoa1211776
32 Tonelli M, Wiebe N, Fortin M, Guthrie B, Hemmelgarn BR, James MT, et al. Methods for identifying 30 chronic conditions: application to administrative data. BMC Med Inform Decis Mak. 2015;15(1):31. https://doi.org/10.1186/s12911-015-0155-5
33 Tammemägi MC, Ruparel M, Tremblay A, Myers R, Mayo J, Yee J, et al. USPSTF2013 versus PLCOm2012 lung cancer screening eligibility criteria (International Lung Screening Trial): interim analysis of a prospective cohort study. Lancet Oncol. 2022;23(1):138–48. https://doi.org/10.1016/s1470-2045(21)00590-8
34 Lin DY, Ying Z. Semiparametric analysis of general additive-multiplicative hazard models for counting processes. Ann Stat. 1995;23(5):1712–34. https://doi.org/10.1214/aos/1176324320
35 Korn EL, Graubard BI, Midthune D. Time-to-Event analysis of longitudinal follow-up of a survey: choice of the time-scale. Am J Epidemiol. 1997;145(1):72–80. https://doi.org/10.1093/oxfordjournals.aje.a009034
36 Fleming TR, Harrington DP. Counting processes and survival analysis. Hoboken, N.J: Wiley-Interscience; 2005.
37 Klein JP, Van Houwelingen HC, Ibrahim JG, Scheike TH, editors. Handbook of survival analysis. 0 ed. Chapman and Hall/CRC; 2016. https://doi.org/10.1201/b16248
38 Lin DY, Ying Z. Semiparametric analysis of the additive risk model. Biometrika. 1994;81(1):61–71. https://doi.org/10.1093/biomet/81.1.61
39 King M. Statistics for process control engineers: a practical approach. 1st ed. Hoboken, NJ: Wiley; 2018.
40 Ahsanullah M, Shakil M, Kibria BG. On a generalized raised cosine distribution: some properties, characterizations and applications. Moroc J Pure Appl Anal. 2019;5(1):63–85. https://doi.org/10.2478/mjpaa-2019-0006
41 Cox DR, Hinkley DV. Theoretical statistics. 0 ed. Chapman and Hall/CRC; 1979. https://doi.org/10.1201/b14832
42 Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics. 2015;31(21):3555–7. https://doi.org/10.1093/bioinformatics/btv402
43 McAvoy S, Ganapathiraju SC, Ducharme-Smith AL, Pritchett JR, Kosari F, Perez DS, et al. Non-random inactivation of large common fragile site genes in different cancers. Cytogenet Genome Res. 2007;118(2–4):260–9. https://doi.org/10.1159/000108309
44 Diaz-Barreiro A, Huard A, Palmer G. Multifaceted roles of IL-38 in inflammation and cancer. Cytokine. 2022;151:155808. https://doi.org/10.1016/j.cyto.2022.155808
45 Xie L, Huang Z, Li H, Liu X, Zheng S, Su W. IL-38: a new player in inflammatory autoimmune disorders. Biomolecules. 2019;9(8):345. https://doi.org/10.3390/biom9080345
46 Marques CR, Costa GN, da Silva TM, Oliveira P, Cruz AA, Alcantara-Neves NM, et al. Suggestive association between variants in IL1RAPL and asthma symptoms in Latin American children. Eur J Hum Genet. 2017;25(4):439–45. https://doi.org/10.1038/ejhg.2016.197
47 Tominaga M, Okamoto M, Kawayama T, Matsuoka M, Kaieda S, Sakazaki Y, et al. Overexpression of IL-38 protein in anticancer drug-induced lung injury and acute exacerbation of idiopathic pulmonary fibrosis. Respir Investig. 2017;55(5):293–9. https://doi.org/10.1016/j.resinv.2017.06.001