



**HAL**  
open science

## A Case-Only Genome-Wide Interaction Study of Smoking and Bladder Cancer Risk: Results from the COBLAnCE Cohort

Maryam Karimi, Sebastian Mendez-Pineda, H len  Blanch , Anne Boland, C line Besse, Jean Fran ois Deleuze, Xiangyu Meng, Nanor Sirab, Karine Groussard, Thierry Lebret, et al.

► **To cite this version:**

Maryam Karimi, Sebastian Mendez-Pineda, H len  Blanch , Anne Boland, C line Besse, et al.. A Case-Only Genome-Wide Interaction Study of Smoking and Bladder Cancer Risk: Results from the COBLAnCE Cohort. *Cancers*, 2023, 15 (17), 10.3390/cancers15174218 . hal-04229518

**HAL Id: hal-04229518**

**<https://hal.science/hal-04229518>**

Submitted on 12 Oct 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin e au d p t et   la diffusion de documents scientifiques de niveau recherche, publi s ou non,  manant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv s.



Distributed under a Creative Commons Attribution 4.0 International License

## Article

# A Case-Only Genome-Wide Interaction Study of Smoking and Bladder Cancer Risk: Results from the COBLAnCE Cohort

Maryam Karimi <sup>1,2,\*</sup> , Sebastian Mendez-Pineda <sup>1</sup>, H el ene Blanch e <sup>3</sup>, Anne Boland <sup>4</sup>, C eline Besse <sup>4</sup> , Jean-Fran ois Deleuze <sup>3,4</sup>, Xiang-Yu Meng <sup>5</sup>, Nanor Sirab <sup>6</sup>, Karine Groussard <sup>1</sup>, Thierry Leuret <sup>7</sup>, Julia Bonastre <sup>1,2</sup>, Yves Allory <sup>5,8</sup>, Fran ois Radvanyi <sup>5</sup>, Simone Benhamou <sup>1,†</sup>  and Stefan Michiels <sup>1,2,†</sup> 

<sup>1</sup> Oncostat U1018 Inserm,  quipe Labellis e Ligue Contre le Cancer, Universit  Paris-Saclay, 94805 Villejuif, France

<sup>2</sup> Bureau de Biostatistique et d' pid miologie, Gustave Roussy, Universit  Paris-Saclay, 94805 Villejuif, France

<sup>3</sup> CEPH-Biobank, Fondation Jean Dausset-CEPH, 75010 Paris, France

<sup>4</sup> Centre National de Recherche en G nomique Humaine (CNRGH), CEA, Universit  Paris-Saclay, 91057 Evry, France

<sup>5</sup> CNRS UMR144, Curie Institute, 75005 Paris, France

<sup>6</sup> Curie Institute, CNRS, UMR144, Molecular Oncology Team, PSL Research University, 75005 Paris, France

<sup>7</sup> Urology, Foch Hospital, 92150 Suresnes, France

<sup>8</sup> UVSQ, Curie Institute, Department of Pathology, Universit  Paris-Saclay, 92210 Saint-Cloud, France

\* Correspondence: maryam.karimi@gustaveroussy.fr

† These authors contributed equally to this work.

**Simple Summary:** Bladder cancer is one of the most common cancers worldwide. The most important known cause of bladder cancer is tobacco smoking. Several genetic variations are also linked to the risk of developing bladder cancer. In this study, we focused on how smoking behavior interacts with genes to affect bladder cancer risk using the French national prospective COBLAnCE cohort (COhort to study BLadder CancEr). The COBLAnCE cohort comprises approximately 1700 bladder cancer patients with available smoking and genetic information. We identified new chromosomal regions (specifically, genes located at 4q22.1, 12p13.1, and 16p13.3) of which variations may interact with smoking behavior in relation to bladder cancer risk. These findings need to be replicated in other studies.

**Abstract:** Bladder cancer (BC) is the 6th most common cancer worldwide, with tobacco smoking considered as its main risk factor. Accumulating evidence has found associations between genetic variants and the risk of BC. Candidate gene-environment interaction studies have suggested interactions between cigarette smoking and *NAT2/GSTM1* gene variants. Our objective was to perform a genome-wide association case-only study using the French national prospective COBLAnCE cohort (COhort to study BLadder CancEr), focusing on smoking behavior. The COBLAnCE cohort comprises 1800 BC patients enrolled between 2012 and 2018. Peripheral blood samples collected at enrolment were genotyped using the Illumina Global Screening Array with a Multi-Disease drop-in panel. Genotyping data (9,719,614 single nucleotide polymorphisms (SNP)) of 1674, 1283, and 1342 patients were analyzed for smoking status, average tobacco consumption, and age at smoking initiation, respectively. A genome-wide association study (GWAS) was conducted adjusting for gender, age, and genetic principal components. The results suggest new candidate loci (4q22.1, 12p13.1, 16p13.3) interacting with smoking behavior for the risk of BC. Our results need to be validated in other case-control or cohort studies.

**Keywords:** bladder cancer; genome-wide association study; smoking; case-only design



**Citation:** Karimi, M.; Mendez-Pineda, S.; Blanch e, H.; Boland, A.; Besse, C.; Deleuze, J.-F.; Meng, X.-Y.; Sirab, N.; Groussard, K.; Leuret, T.; et al. A Case-Only Genome-Wide Interaction Study of Smoking and Bladder Cancer Risk: Results from the COBLAnCE Cohort. *Cancers* **2023**, *15*, 4218. <https://doi.org/10.3390/cancers15174218>

Academic Editor: Roman Blaheta

Received: 6 July 2023

Revised: 2 August 2023

Accepted: 7 August 2023

Published: 23 August 2023



**Copyright:**   2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Bladder cancer (BC) ranked tenth among the most common cancer types worldwide, with an estimated total of 573,278 new bladder cancer cases and 212,536 deaths due to bladder cancer in 2020 [1]. Over three-fourths of all BC cases occur in men [1].

Bladder cancer risk factors can be divided into external exposure and genetic predisposition variants. Studies have suggested an increased risk of bladder cancer associated with occupational exposure to carcinogens [2–4] and water contaminants [5]. However, the most important known bladder cancer risk factor is tobacco smoking [6–8]. It is estimated that almost 50% of bladder cancer cases are attributable to smoking [9] and can be prevented.

Previous studies suggested a potential role of genetic predisposition in bladder cancer occurrence [10]. Recent genome-wide association studies (GWAS) have been successful in identifying variants associated with bladder cancer [11–15]. These studies identified susceptibility loci at 3q28 (variants with the nearest gene *TP63*) [14], 4p16.3 (variants with the nearest gene *TACC3*) [13], and 8q24.21 (variants with the nearest gene *MYC*) [14]. However, most of these variants are associated with a small bladder cancer risk. To provide more insights into the biological mechanism of bladder cancer, gene-environment (G×E) interactions have been proposed. Interactions between cigarette smoking (as the major risk of bladder cancer) and N-acetyl transferase 2 (*NAT2*) and Glutathione-s-transferaseM1 (*GSTM1*) variants, the two best-known candidates, have been consistently associated with BC development [16]. Indeed, studies have shown a higher relative risk of bladder cancer due to smoking for *NAT2* slow acetylators compared to *NAT2* rapid/intermediate acetylators. This was justified by the fact that *NAT2* detoxifies aromatic amines (primary bladder cancer carcinogen in tobacco) [17], and hence the effect of smoking becomes limited to smokers. The G×E interaction studies are important as they help to identify potentially new loci interacting with environmental exposure that would not be revealed by the main GWAS analysis. The results of such studies can help to set up new strategies for bladder cancer prevention and control. To date, most G×E interaction studies have focused on candidate genes, such as carcinogen detoxification genes, and much less attention has been given to ‘agnostic’, free-of-candidate, G×E interactions. Two studies went further and performed these association analyses on a wider range of single nucleotide polymorphisms (SNPs) (using more than 490,000 genotyped data [18] or imputed SNPs [19]), and found some suggestive evidence of interactions (either additive or multiplicative) between smoking and new SNPs and bladder cancer risk [18], but there is still a need for more data on the potential interaction between genetic variants and smoking habits.

However, conducting a G×E interaction study could be challenging due to the size of the study and the limited power for interaction analysis. The standard approach to test the interaction when data comes from case-control studies is to use a regression model by adding an interaction term between the variables of interest (for example, genetic variants and smoking phenotypes). Using this approach to detect an interaction effect of the same magnitude as the main effect (obtained from GWAS analysis), the sample size should be at least four times the sample used for the main effects analysis [20]. In the absence of a control group, the case-only design has been proposed as an alternative approach with more power for interaction tests [21,22].

Here, we aimed to explore the genome-wide gene-smoking interactions in bladder cancer development. While previous studies have mostly focused on smoking status and tobacco consumption [18,23–25], we looked at three main smoking phenotypes that characterize individuals’ smoking behavior throughout their life: age at smoking initiation, smoking status (never vs. ever smokers), and average tobacco consumption. We used the COBLAnCE cohort (COhort to study BLadder CancEr) [26], a large prospective French-based cohort containing bladder cancer patients for whom a wide range of phenotypic and environmental measures and biological samples were prospectively collected. We conducted a genome-wide gene-smoking interaction study in bladder cancer based on ~9,700,000 imputed SNPs data from ~1700 bladder cancer patients of the European ancestry population, using a case-only design framework.

## 2. Materials and Methods

### 2.1. Study Population

COBLAnCE is a multicentre cohort of bladder cancer patients (eleven public and three private hospitals). It includes newly diagnosed patients with bladder cancer who were within 1 year of their initial diagnosis and aged 18 years or older. Recruitment was performed from December 2012 to June 2018. A total of 1800 patients were enrolled in the COBLAnCE cohort. Demographic information and detailed smoking data were collected from questionnaires completed by trained nurses who asked questions of patients at baseline.

### 2.2. Smoking Phenotypes

A complete history of tobacco smoking was ascertained separately for cigarettes, cigars, and pipes. Depending on the availability of data in the COBLAnCE cohort and based on previous genetic associations in the literature, three phenotypes were selected:

Smoking status: whether the patient reported as never smoker (less than 100 cigarettes/lifetime) or ever smoker. This phenotype was considered a binary phenotype coded by 0 (never smoker) or 1 (ever smoker).

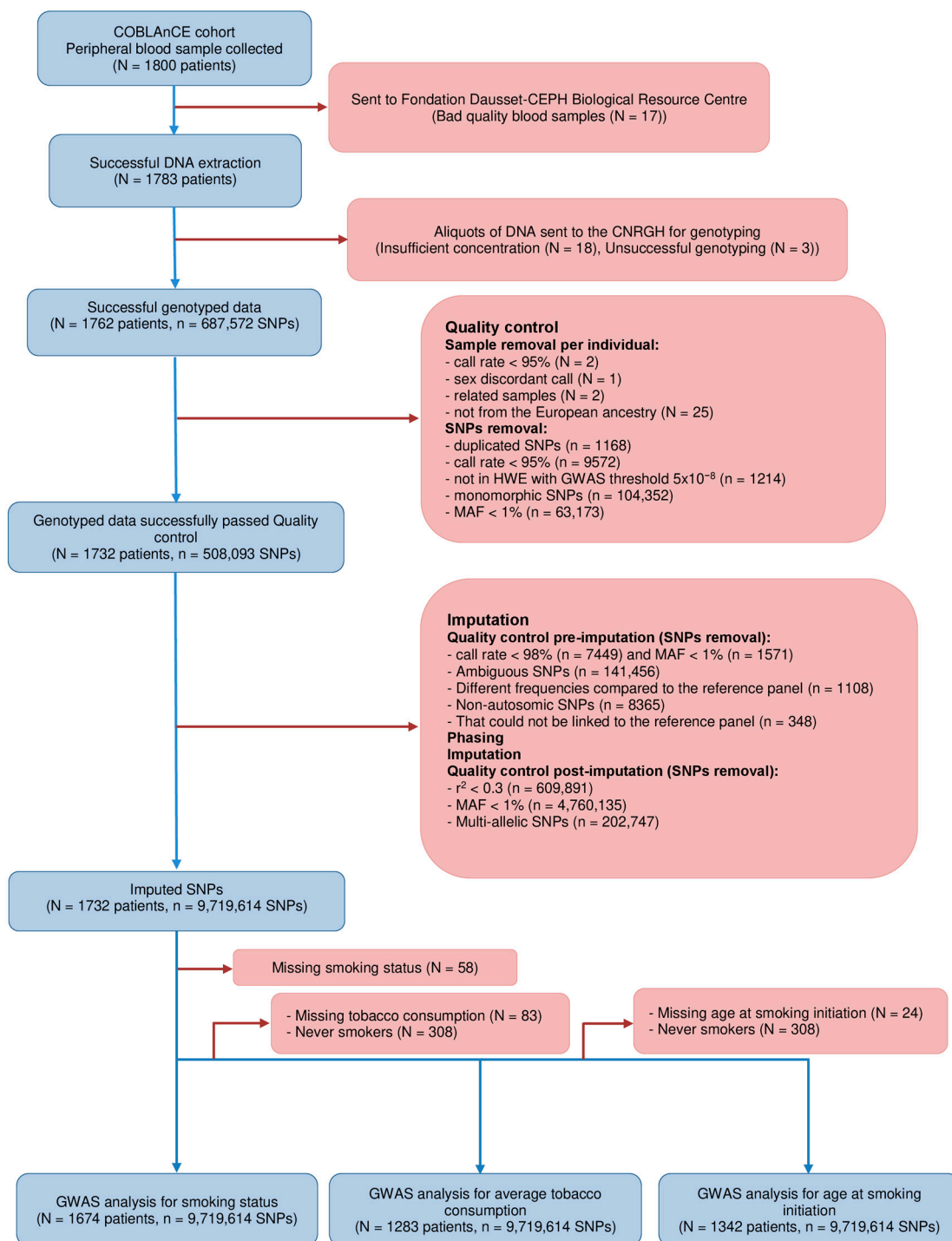
Age at smoking initiation: the age at which the patient started smoking (cigarette/pipe or cigar). This phenotype was considered a continuous phenotype and was log-transformed since it is highly skewed.

Average tobacco consumption in grams for ever-smokers: using a complete history of tobacco smoking collected separately for cigarettes, cigars, and pipes. To standardize consumption across products, we calculated product-specific lifetime consumption (in grams of tobacco) based on the International Agency for Research on Cancer estimates of average unit weights [27,28]. The average tobacco consumption was then calculated by dividing the cumulative lifetime tobacco consumption (in grams) by the overall duration of smoking and was log-transformed since this variable is highly skewed. Given its construction, this phenotype combines both tobacco consumption and the duration of smoking across a lifetime.

The analyses of the phenotypes of age at smoking initiation and average tobacco consumption excluded never smokers.

### 2.3. Genotype Data, Quality Control, and Imputation

Peripheral blood samples were collected from all patients at enrolment (N = 1800). Samples were sent to the CEPH-Biobank, on average, within 48 h after collection for processing. DNA was extracted from buffy-coats using the salting-out method on the Autopure (Qiagen, Hilden, Germany) automated system and quantified using fluorimetry with the Quant-iT DNA Assay kit, Broad Range (Thermo Fisher Scientific, Waltham, MA, USA). Aliquots of DNA from 1765 patients were sent to the National Centre of Research in Human Genomics (CNRRGH, CEA, Evry, France) for genotyping using the Illumina Global Screening Array with a Multi-Disease drop-in panel (GSA-MD v1.0, Illumina San Diego, CA, USA), which allows the analysis of ~700,000 polymorphisms. Genotyping was performed at the CNRRGH on an Illumina automated high-throughput genotyping platform, according to the manufacturer's instructions; 2 internal positive controls were included on each plate. All aliquoted samples were genotyped, and genotypes of 1762 patients were transferred for further analysis (failed genotyping for 21 samples, Figure 1).



**Figure 1.** Flowchart of the study population from data collection, genotyping, and imputation to data analyses in the COBLAnCE cohort. GWAS, genome-wide association study. COBLAnCE, COhort to study BLadder CancEr. HWE, Hardy-Weinberg Equilibrium. MAF, minor allele frequency. SNP, single nucleotide polymorphisms.

Quality control (QC) on genotyped data was performed using PLINK version 1.9 [29,30]. We excluded samples with a call rate <95% (N = 2) and with a sex-discordant call (N = 1). Additionally, 2 pairs of related subjects were identified, and for each pair, the patient with the greater number of missing genotype data was excluded. Of the remaining 1757 patients, 1732 were of European ancestry.



Of the 687,572 SNPs that were successfully obtained, we excluded duplicated SNPs ( $n = 1168$ ) with a call rate  $<95\%$  ( $n = 9572$ ), SNPs not in Hardy-Weinberg Equilibrium (HWE) (with the GWAS threshold  $p < 5 \times 10^{-8}$ ) ( $n = 1214$ ), monomorphic SNPs ( $n = 104,352$ ) and SNPs with a minor allele frequency (MAF)  $<1\%$  ( $n = 63,173$ ). The final set of SNPs retained comprised 508,093 SNPs.

Imputation of missing SNPs was performed only on the European ancestry population ( $N = 1732$  patients and  $n = 508,093$  SNPs), and 1000 Genome Phase 3 was used as the reference panel [31]. Additional filtering was applied to the genotyped data to increase the accuracy of the imputation. We, therefore, excluded the following SNPs from imputation: SNPs with a call rate  $<98\%$  ( $n = 7449$ ) and MAF  $<1\%$  ( $n = 1571$ ), ambiguous SNPs ( $n = 141,456$ ), SNPs with different frequencies compared to the reference panel ( $n = 1108$ ), non-autosomal SNPs ( $n = 8365$ ), and SNPs that could not be linked to the reference panel ( $n = 348$ ). Therefore, imputation was based on 347,796 genotyped SNPs. A two-stage approach was considered for the imputation: first, we phased GWAS samples with SHAPEIT (v2.r904) [32] using burn, prune, and main parameters (--burn 10 --prune 10 --main 50) and then genotype imputation using IMPUTE2 [33,34]. We considered non-overlapping intervals defined by the centromere locations of the human reference sequence assembly (GRCh37). For the imputation, we set the buffer region to 500 kb and the number of reference haplotypes to 800. The imputation gave us a sample of 15,292,387 SNPs. Once imputed, we excluded imputed variants with imputation info values ( $r^2$ )  $<0.3$  ( $n = 609,891$ ), MAF  $<1\%$  ( $n = 4,760,135$ ), and multi-allelic SNPs ( $n = 202,747$ ). Our final SNP population was comprised of 9,719,614 SNPs ready for use in our association analyses.

The whole procedure (from blood collection to genotyping and imputation) is outlined in Figure 1.

#### 2.4. Statistical Analysis

GWAS was performed on COBLAnCE cohort participants of European ancestry with available genotyping data after QC and imputation. We employed a case-only study design to estimate gene-environment interactions to test the association between SNPs and smoking phenotypes in bladder cancer patients. The analyses were conducted using logistic regression (for smoking status phenotype) and linear regression models (for age at smoking initiation and average tobacco consumption), with phenotype as the dependent variable against each of the imputed SNPs as the independent variable. Regressions were adjusted for gender, age, and the first ten genetic principal components to control for the study population's latent genetic structure. The results were presented by Manhattan plots and quantile-quantile (Q-Q) plots, and the significance threshold was set at the established GWAS level ( $5 \times 10^{-8}$ ). Functional annotation was performed for significant SNPs and the top hits using ANNOVAR [35] to identify the overlapping or closest gene to each SNP. For these SNPs, effect size estimates were presented as odds ratios (ORs), 95% confidence interval (CI) or  $\beta$ s, standard error (SE), and  $p$ -values. An OR greater than 1.0 and a  $\beta$  greater than 0 suggest a positive interaction for patients with reference alleles compared to those with alternative alleles. The case-only design approach relies on the independence assumption of genetic variants and environmental exposure. If the independence assumption is violated, the number of false positives will increase. A two-step approach was proposed in the literature to overcome the power insufficiency of the case-control approach [36]. In the first step, a case-only design of GxE interaction is performed, and a subset of SNPs is selected. In the second step, the selected SNPs are used in a case-control approach for the GxE interaction. This two-step strategy could help reduce false positive results. Since no control group was available in the COBLAnCE cohort, we opted for a different second step in which we considered a set of 433 SNPs from the literature. The significant association between these SNPs and different smoking phenotypes was shown within a large study of up to 1.2 million individuals [37]. We verified whether one of our top SNPs (10 SNPs with the smallest  $p$ -values) was part of the selected SNPs or was in high linkage disequilibrium (LD) with them.

We also searched the literature for candidate SNPs that reported a smoking interaction association for bladder cancer risk (selected either from association studies with smoking phenotypes in the case-only population or from gene-smoking interaction studies in the case-control population). We narrowed our research to articles published from 2008 onwards. SNPs were filtered and retained with a reported  $p < 10^{-4}$  within the European population and were not associated with smoking, which led us to a total number of 37 SNPs [16,18]. We investigated the corresponding  $p$ -values of these SNPs in our GWAS results.

All statistical analyses were performed using PLINK v1.9 and R version 4.1.1 (<https://www.r-project.org/> (accessed on 6 July 2023)).

### 3. Results

The baseline characteristics of the study population are presented in Table 1. The study population comprised 1732 patients with European ancestry (303 females and 1429 males) with a mean age of 68 years old. The difference in gender ratio was mostly attributable to the higher prevalence of smoking among males than among women in France several years ago. Accordingly, the results for smoking exposure were presented separately for males and females.

**Table 1.** Descriptive characteristics of the study population.

	Overall (N = 1732)	Women (N = 303)	Men (N = 1429)
<b>Age (Years)</b>			
Mean (SD)	68.55 (10.79)	68.05 (12.46)	68.65 (10.41)
Median (Q1, Q3)	68.70 (61.74, 76.34)	68.25 (60.27, 77.31)	68.73 (62.07, 76.10)
Min; Max	22.05; 95.31	22.2; 93.34	22.05; 95.31
Missing	1	0	1
<b>Smoking status</b>			
Never smoker	308 (17.78%)	115 (37.95%)	193 (13.51%)
Ever smoker	1366 (78.87%)	176 (58.09%)	1190 (83.28%)
Missing	58 (3.35%)	12 (3.96%)	46 (3.22%)
<b>Duration of smoking (Years) (N = 1366)</b>			
Mean (SD)	34.10 (14.54)	33.09 (14.17)	34.25 (14.59)
Median (Q1, Q3)	35.00 (24.00, 45.00)	35.00 (22.00, 44.00)	35.00 (24.00, 45.00)
Min; Max	1; 76	1; 62	1; 76
Missing	83	17	66
<b>Average tobacco consumption in grams (N = 1366)</b>			
Mean (SD)	19.82 (12.65)	16.00 (11.45)	20.36 (12.72)
Median (Q1, Q3)	18.61 (11.43, 23.51)	14.41 (9.23, 20.00)	19.11 (12.16, 24.67)
Min; Max	0.5; 115	1; 62.06	0.5; 115
Missing	83	17	66
<b>Age at smoking initiation (Years) (N = 1366)</b>			
Mean (SD)	18.40 (5.83)	20.79 (7.46)	18.06 (5.47)
Median (Q1, Q3)	18.00 (15.00, 20.00)	18 (17.00, 22.00)	17 (15.00, 20.00)
Min; Max	7; 73	10; 56	7; 73
Missing	24	5	19

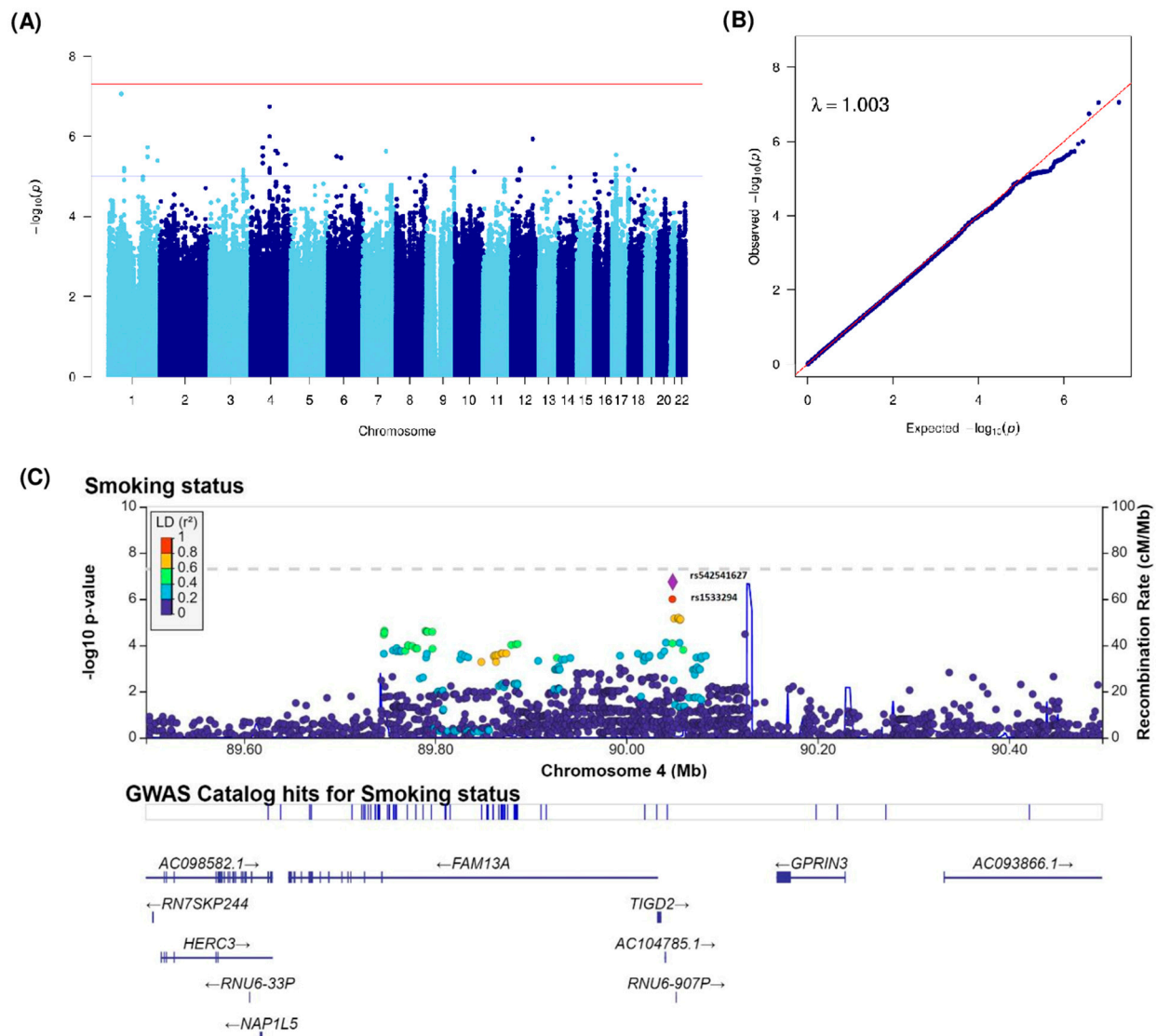
SD, standard deviation; Q1, first quartile; Q3, third quartile.

Of these 1732 patients, 308 were never smokers, and 1366 were former smokers with missing smoking status for 58 patients (3.35%). The proportion of never smokers was higher among women (37.5%) than men (13.51%).

The mean duration of tobacco consumption was almost the same for men and women (33 years vs. 34 years, respectively). However, men consumed more tobacco than women (19 g on average vs. 14 g, respectively). It seems that men started smoking earlier than women (18 years old vs. 20 years old, respectively).

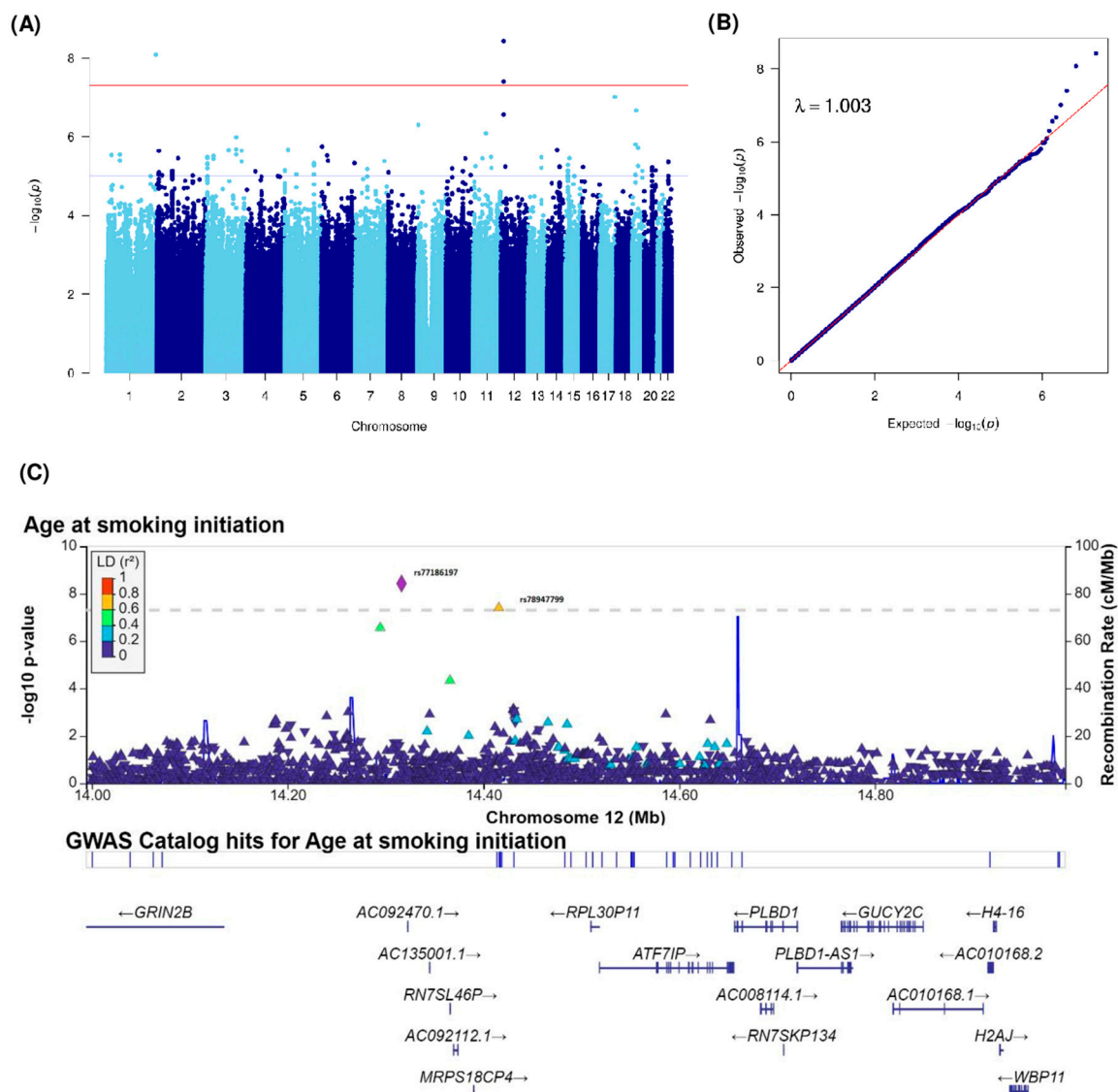
As shown in Figure 1, analyses for smoking status were carried out on 1674 patients out of 1732 patients because of the missing smoking status of 58 patients. For the age at smoking initiation and average tobacco consumption, analyses were restricted to 1342 patients

and 1283 patients, respectively, because of the missing age at smoking initiation for 24 patients, having never smoker status for 308 patients, and missing tobacco consumption for 83 patients. We performed a GWAS using a case-only design. The Manhattan plot of  $-\log_{10}(p)$  and the Q-Q plots of the observed  $p$ -values versus the expected  $p$ -values and the genomic inflation factor ( $\lambda$ ) of our GWAS analyses for smoking status, age at smoking initiation, and average tobacco consumption phenotypes are shown in Figure 2A,B, Figure 3A,B and Figure 4A,B. No genomic inflation was detected in our GWAS ( $\lambda = 1.003, 1.003, \text{ and } 1.002$  for analyses of smoking status, age at smoking initiation, and average tobacco consumption, respectively).



**Figure 2.** (A) Manhattan and (B) Q-Q plots of GWAS for interactions with smoking status. Blue and red horizontal lines indicate  $1 \times 10^{-5}$  and  $5 \times 10^{-8}$  (the typical GWAS threshold) thresholds, respectively. (C) Regional association plot for the 4q22.1 locus interacting with smoking status. The most significantly associated SNP in the region is presented in a diamond (rs542541627). Other dots correspond to other analyzed SNPs. Dots are colored by their linkage disequilibrium coefficient. GWAS, genome-wide association study. Q-Q plots, quantile–quantile plots. SNP, single nucleotide polymorphisms.





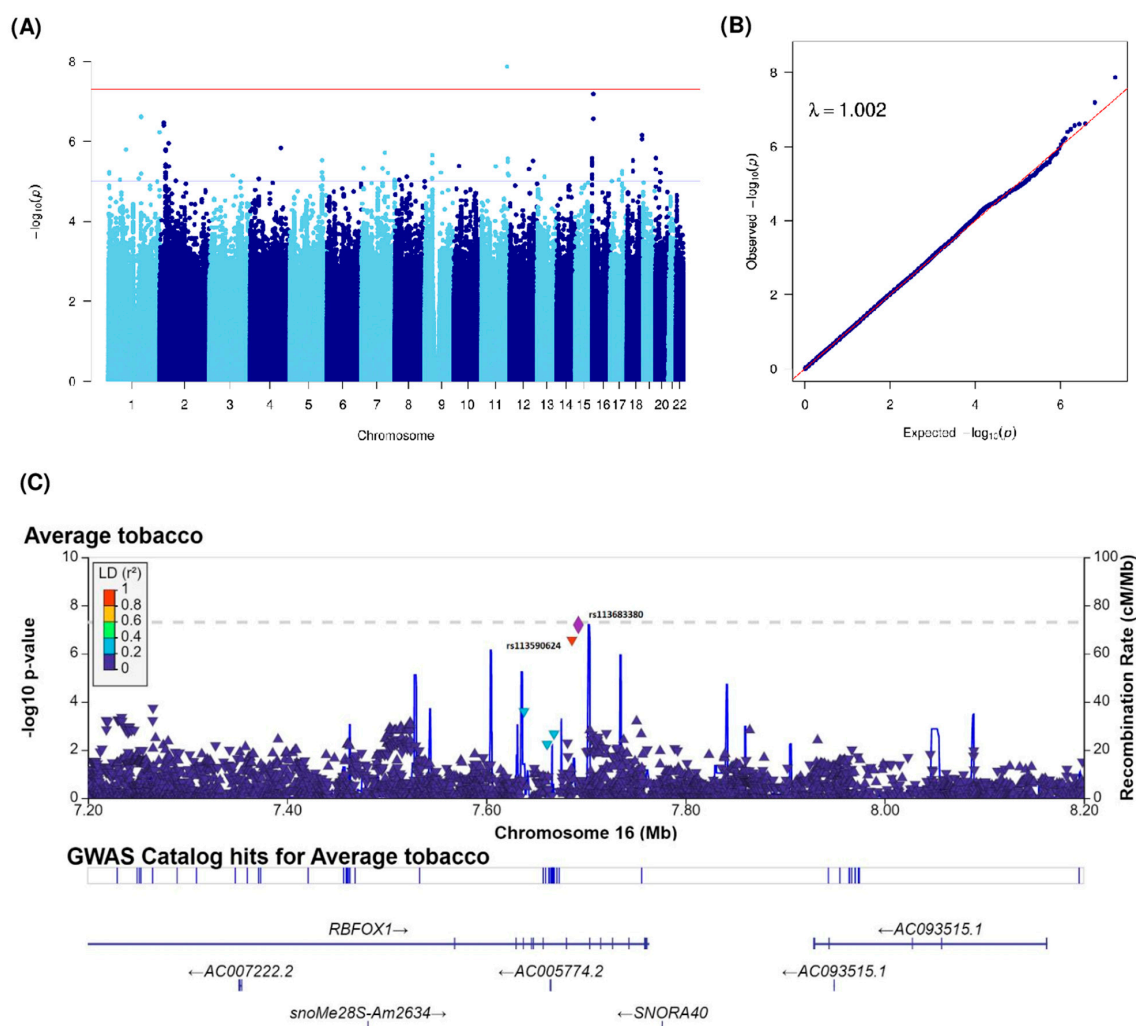
**Figure 3.** (A) Manhattan and (B) Q-Q plots of GWAS for interaction with age at smoking initiation. Blue and red horizontal lines indicate  $1 \times 10^{-5}$  and  $5 \times 10^{-8}$  (the typical GWAS threshold) thresholds, respectively. (C) Regional association plot for the 12p13.1 locus interacting with age at smoking initiation. The most significantly associated SNP in the region is presented in a diamond (rs77186197). Other dots correspond to other analyzed SNPs. Dots are colored by their linkage disequilibrium coefficient. GWAS, genome-wide association study. Q-Q plots, quantile–quantile plots. SNP, single nucleotide polymorphisms.

The top ten SNPs for each GWAS are presented in Table 2, for which only four SNPs reached the genome-wide significance threshold ( $p < 5 \times 10^{-8}$ ). None of these SNPs were mentioned previously in the literature in association with smoking phenotypes and were not in high LD with the 433 selected SNPs (squared correlation  $r_{LD}^2$  between pair of SNPs  $\leq 0.90$ , Supplementary Figure S1). Therefore, we presumed the independence of SNPs and smoking, which is the main hypothesis of a case-only design.

For the smoking status phenotype, none of the SNPs reached the GWAS significance threshold. However, the most significant SNP variants interacting with smoking status for BC risk were rs114073636 and rs116571608, located on chromosome 1p31.3 (OR = 0.31, 95% CI = [0.20–0.47],  $p = 8.68 \times 10^{-8}$  and OR = 0.31, 95% CI = [0.20–0.47],  $p = 8.87 \times 10^{-8}$ , respectively for patients with reference allele compared to those with alternative allele). Both variants were intergenic. Our results also suggest a promising locus on 4q22.1 (not

reaching the GWAS threshold but significant at the  $10^{-5}$  threshold) with an intergenic SNP located between genes *TIGD2* and *GPRIN3* (Figure 2A). The regional association plot of this locus showed high LD with surrounding SNPs (Figure 2C).

Analyses yielded three SNPs significantly interacting with age at smoking initiation (Figure 3A). The reference allele at rs531756449, an intergenic variant located on chromosome 1q44, interacted positively with age at smoking initiation ( $\beta(\text{SE}) = 0.3 (0.05)$ ,  $p = 8.26 \times 10^{-9}$ ). The reference allele at SNPs located on the 12p13.1 locus showed positive interaction associations with age at smoking initiation and bladder cancer risk (rs77186197 with  $\beta(\text{SE}) = 0.29 (0.05)$ ,  $p = 3.74 \times 10^{-9}$ , rs78947799 with  $\beta(\text{SE}) = 0.31 (0.06)$ ,  $p = 3.97 \times 10^{-8}$  and rs79782126 with  $\beta(\text{SE}) = 0.18 (0.03)$ ,  $p = 2.76 \times 10^{-7}$ ). The regional association plot of the 12p13.1 locus (Figure 3C) mapped between genes *GRIN2B* and *ATF7IP* shows elevated LD with two other surrounding variants ( $>0.4$ ) that did not reach the significance threshold.



**Figure 4.** (A) Manhattan and (B) Q-Q plots of GWAS for the interaction with average tobacco consumption. Blue and red horizontal lines indicate  $1 \times 10^{-5}$  and  $5 \times 10^{-8}$  thresholds (the typical GWAS threshold), respectively; (C) Regional association plot for the 16p13.3 locus interacting with average tobacco consumption. The most significantly associated SNP in the region is presented in a diamond (rs113683380). Other dots correspond to other analyzed SNPs. Dots are colored by their linkage disequilibrium coefficient. GWAS, genome-wide association study. Q-Q plots, quantile-quantile plots. SNP, single nucleotide polymorphisms.

**Table 2.** Top 10 SNPs of GWAS for interaction with smoking phenotypes in the COBLAnCE cohort.

Phenotype	rsID	Chr	Position	Locus	REF	ALT	MAF	OR (95% CI)	β (SE)	p-value	Annotation (Distance)	Type
Smoking status (Ever vs. Never)	rs114073636	1	62066340	1p31.3	G	A	0.04	0.31 (0.20–0.47)		$8.68 \times 10^{-8}$	NFLA (137880), MGC34796 (53574)	intergenic
	rs116571608	1	62062694	1p31.3	G	A	0.04	0.31 (0.20–0.47)		$8.87 \times 10^{-8}$	NFLA (134234), MGC34796 (57220)	intergenic
	rs2110040	1	187842047	1q31.1	G	C	0.01	0.17 (0.08–0.36)		$1.88 \times 10^{-6}$	LINC01037 (395693)	intergenic
	rs542541627	4	90048466	4q22.1	A	AAAAACAAACAAAC	0.44	1.69 (1.39–2.06)		$1.81 \times 10^{-7}$	TIGD2 (12414), GPRIN3 (109068)	intergenic
	rs1533294	4	90048122	4q22.1	C	T	0.42	1.63 (1.34–1.98)		$1.01 \times 10^{-6}$	TIGD2 (12070), GPRIN3 (109412)	intergenic
	rs80281369	4	55856707	4q12	T	C	0.04	0.36 (0.24–0.55)		$1.91 \times 10^{-6}$	KIT (249826), KDR (87941)	intergenic
	rs11098419	4	118939953	4q26	T	G	0.30	1.62 (1.33–1.98)		$2.30 \times 10^{-6}$	LINC02264 (148850), NDST3 (15688)	intergenic
	rs115317515	4	129223251	4q28.2	C	T	0.02	0.29 (0.17–0.49)		$2.64 \times 10^{-6}$	PGRMC2 (14283), LINC02615 (125920)	intergenic
	rs76261406	7	110492301	7q31.1	T	G	0.03	0.23 (0.13–0.42)		$2.39 \times 10^{-6}$	IMMP2L	intronic
Age at smoking initiation (Years)	rs11112182	12	105139116	12q23.3	A	G	0.19	0.57 (0.45–0.71)		$1.16 \times 10^{-6}$	CHST11	intronic
	<b>rs531756449</b>	<b>1</b>	<b>244460577</b>	<b>1q44</b>	<b>C</b>	<b>G</b>	<b>0.02</b>		<b>0.3 (0.05)</b>	<b><math>8.26 \times 10^{-9}</math></b>	<b>ZBTB18 (239797), C1orf100 (55360)</b>	<b>intergenic</b>
	rs140538571	3	149654460	3q25.1	C	A	0.01		−0.24 (0.05)	$1.04 \times 10^{-6}$	RNF13	intronic
	rs115421081	3	149600807	3q25.1	C	G	0.01		−0.24 (0.05)	$1.05 \times 10^{-6}$	RNF13	intronic
	rs148961658	9	4111766	9p24.2	G	C	0.01		0.26 (0.05)	$5.03 \times 10^{-7}$	GLIS3	intronic
	rs117818261	11	62644214	11q12.3	C	T	0.01		0.21 (0.04)	$8.13 \times 10^{-7}$	SLC3A2	intronic
	<b>rs77186197</b>	<b>12</b>	<b>14316383</b>	<b>12p13.1</b>	<b>C</b>	<b>T</b>	<b>0.01</b>			<b><math>3.74 \times 10^{-9}</math></b>	<b>GRIN2B (183093), ATF7IP(202183)</b>	<b>intergenic</b>
	<b>rs78947799</b>	<b>12</b>	<b>14415811</b>	<b>12p13.1</b>	<b>A</b>	<b>C</b>	<b>0.01</b>			<b><math>3.97 \times 10^{-8}</math></b>	<b>GRIN2B (282521), ATF7IP (102755)</b>	<b>intergenic</b>
	rs79782126	12	14294621	12p13.1	T	C	0.02		0.18 (0.03)	$2.76 \times 10^{-7}$	GRIN2B (161331), ATF7IP (223945)	intergenic
	rs149790626	17	72716530	17q25.1	C	T	0.01		0.35 (0.06)	$9.76 \times 10^{-8}$	RAB37	intronic
rs34177209	19	18474978	19p13.11	T	A	0.26		0.06 (0.01)	$1.57 \times 10^{-6}$	PGPEP1	UTR3	

Table 2. Cont.

Phenotype	rsID	Chr	Position	Locus	REF	ALT	MAF	OR (95% CI)	$\beta$ (SE)	p-value	Annotation (Distance)	Type
Average tobacco consumption (grams)	rs114681930	1	160046712	1q23.2	G	T	0.01		−0.7 (0.14)	$2.39 \times 10^{-7}$	KCNJ10 (6762), KCNJ9 (4616)	intergenic
	rs79752468	1	160050867	1q23.2	T	A	0.01		−0.7 (0.14)	$2.44 \times 10^{-7}$	KCNJ9 (461)	upstream
	rs142728151	1	248166173	1q44	G	A	0.02		−0.58 (0.12)	$5.94 \times 10^{-7}$	OR2L13 LINC01376	intronic
	rs114634507	2	19247230	2p24.1	G	A	0.02		−0.65 (0.13)	$3.44 \times 10^{-7}$	(20504), MIR4757 (300960) LINC01376	intergenic
	rs149246142	2	19292223	2p24.1	G	A	0.02		−0.67 (0.13)	$4.01 \times 10^{-7}$	(65497), MIR4757 (255967)	intergenic
	<b>rs2714069</b>	<b>11</b>	<b>123389614</b>	<b>11q24.1</b>	<b>A</b>	<b>G</b>	<b>0.03</b>		<b>0.63 (0.11)</b>	<b><math>1.35 \times 10^{-8}</math></b>	<b>GRAMD1B</b>	<b>intronic</b>
	rs113683380	16	7692534	16p13.3	G	A	0.02		−0.66 (0.12)	$6.44 \times 10^{-8}$	RBF0X1	intronic
	rs113590624	16	7686089	16p13.3	G	C	0.02		−0.64 (0.12)	$2.74 \times 10^{-7}$	RBF0X1	intronic
	rs146671367	18	72146124	18q22.3	C	T	0.01		−0.92 (0.18)	$6.89 \times 10^{-7}$	DIPK1C (21621), CNDP2 (17474)	intergenic
	rs151292117	18	72145473	18q22.3	G	A	0.01		−0.94 (0.19)	$8.84 \times 10^{-7}$	DIPK1C (20970), CNDP2 (18125)	intergenic

ALT, alternative allele; CI, confidence interval; COBLAnCE, COhort to study BLadder CancEr; GWAS, genome-wide association study; MAF, minor allele frequency; REF, reference allele; OR, odds ratio; rsID, SNP identification number; SNP, single nucleotide polymorphism. SNPs that are significant at the GWAS threshold are presented in bold.

The reference allele of an intronic SNP, located on gene *GRAMD1B* on chromosome 11q24.1, significantly interacted positively with average tobacco consumption (rs2714069 with  $\beta(\text{SE}) = 0.63 (0.11)$ ,  $p = 1.35 \times 10^{-8}$ ). However, the reference alleles of SNPs located on locus 16p13.3 seems to have protective variants for the interaction between average tobacco consumption (rs113683380 with  $\beta(\text{SE}) = -0.66 (0.12)$ ,  $p = 6.44 \times 10^{-8}$  and rs113590624 with  $\beta(\text{SE}) = -0.64 (0.12)$ ,  $p = 2.74 \times 10^{-7}$ ). These variants were intronic and located on gene *RBFOX1* (Figure 4A). The regional association plot of the 16p13.3 locus showed high LD between these two promising variants (Figure 4C).

We also investigated previously published SNPs in the literature in our GWAS results. Supplementary Table S1 shows a comparison of our results for the 37 selected SNPs. None of the SNPs were close to the GWAS threshold. However, the SNP with the smallest  $p$ -value was rs1495741, which showed a positive interaction with smoking initiation (OR = 1.37, 95% CI = [1.11–1.70],  $p = 0.004$  for patients with reference allele compared to alternative allele), located on *NAT2*.

#### 4. Discussion

The current paper focused on the interaction between SNPs and different smoking phenotypes in relation to bladder cancer risk. To our knowledge, this is the third study that focuses on free-of-candidate GxE interactions.

We performed a genome-wide interaction study of smoking phenotypes and BC risk within a case-only design framework. Our analyses were based on the COBLAnCE study, which to our knowledge, is one of the largest prospective cohorts of bladder cancer patients with numerous data and sequential biological samples collected at baseline and during the follow-up. Such prospective cohorts with large sample sizes are rare because of the major efforts required for patient recruitment and data/biomaterial collection.

Within the COBLAnCE cohort, we identified multiple loci on chromosomes 4q22.1, 12p13.1, and 16p13.3 as new candidates containing SNPs interacting with our smoking phenotypes (smoking status, age at smoking initiation and average tobacco smoking, respectively) for the risk of bladder cancer.

The reference allele at SNPs on locus 4q22.1 (rs542541627 and rs1533294) suggested a positive interaction with smoking status. These variants were intergenic and were mapped between genes *TIGD2* and *GPRIN3*. The closest gene, *TIGD2* (Tigger Transposable Element Derived 2), is a protein-coding gene. The other gene, *GPRIN3* (G Protein-Regulated Inducer of Neurite Outgrowth 3), a protein-coding gene, is predicted to be involved in neuron projection development and to be active in the plasma membrane. It has also been suggested that *GPRIN3* is associated with COPD (severe chronic obstructive pulmonary disease) and emphysema [38], and cocaine addiction [39].

The reference allele at our significant SNPs on locus 12p13.1 (rs77186197 and rs78947799) showed positive interactions with age at smoking initiation and was mapped between *GRIN2B* and *ATF7IP*. This suggests that these variants are more relevant in interaction with the early stages of smoking. The closest gene to these variants was *GRIN2B* (Glutamate Ionotropic Receptor NMDA Type Subunit 2B). A protein-coding gene that encodes a member of the N-methyl-D-aspartate (NMDA) receptor family within the ionotropic glutamate receptor superfamily. Its pathways are Transcriptional Regulation by *MECP2* (Methyl-CpG-binding protein 2) (reactome: R-HSA-8986944), which regulates neuronal receptors and channels (reactome: R-HSA-9022699), Disorders of Nervous System Development (reactome: R-HSA-9697154), and Nicotine (KEGG: mpah05033) and cocaine (KEGG: mpah05030) addiction. A new study showed that the overexpression of *MECP2* attenuates cigarette smoke extract-induced lung epithelial cell injury [40]. *GRIN2B* is also associated with nicotine dependence [41] and is one of the genes influencing smoking behaviors [42]. However, the *ATF7IP* (Activating Transcription Factor 7 Interacting Protein), a protein-coding gene, is a multifunctional nuclear protein associated with heterochromatin. *ATF7IP* is inactivated in 5% of lung adenocarcinomas among non-smokers [43].



The reference allele of two of the promising SNPs (rs113683380 and rs113590624, with a negative interaction with average tobacco consumption) located on 16p13.3 was intronic and mapped to the *RBFOX1* gene. *RBFOX1* (RNAs Binding FOX-1 Homolog 1) is a protein-coding gene that has a role in alternative splicing regulation. *RBFOX1* (also known as *A2BP1*) is highly expressed in neurons in brain regions that include the hippocampus, and there is evidence of associations between this gene, smoking cessation, and nicotine dependence [44,45]. Interestingly, one of the pathways of *RBFOX1* is transcriptional regulation by *MECP2* (reactome: R-HSA-8986944). The *RBFOX1* gene was identified as associated with substance dependence or the ability to quit smoking in 13 independent datasets [46], with smoking frequency [47]. In addition, *RBFOX1* was presented as a strong candidate for susceptibility to aggressive behavior and to several psychiatric disorders [48,49]. Our results also showed a significant interaction association with average tobacco consumption for the reference allele of an intronic SNP (rs2714069) located on the *GRAMD1B* (GRAM Domain Containing 1B) gene, a protein-coding gene. A study of breast epithelial cells showed significant upregulation in response to nicotine stress by *GRAMD1B* [50].

It is hard to compare the results of the current study with previously GxE interaction and bladder cancer risk results since most studies are based on case-control cohorts focusing only on smoking status (never vs. ever smokers). Considering a case-only design framework, our results should be interpreted as the multiplicative interaction effect of smoking, genetic variants, and the risk of bladder cancer. On the other side, most studies in the literature presented the additive interaction effect, besides two studies that also looked at the multiplicative interaction effect by performing stratified analyses [18,19]. For example, looking at the multiplicative interaction effect of *NAT2* and smoking status within a meta-analysis study [19], the same direction of the results is obtained within the COBLAnCE cohort, but the *p*-value of *NAT2* in our study did not reach the GWAS significant threshold which could be due to the small sample size compared to a meta-analysis approach.

The strength of our study comes from the COBLAnCE cohort, one of the largest prospective bladder cancer cohorts besides cohorts from Spain [51], the Netherlands [52], the UK [53], and the US [54,55]. Given the structure of the COBLAnCE cohort, it gives an accurate image of practices in bladder cancer patients in France, covering both public and private hospitals. Despite this, one of the limitations of our study is the absence of disease-free individuals. We carried out a case-only design; however, the main assumption of this approach, which is the independence between exposure and genetic variants, could not be validated in this cohort. The second limitation is the absence of replication. However, performing an exact replication study on other bladder cancer cohorts requires the availability of genotyped data and an adequate sample size, which in our case, would be smaller than the COBLAnCE cohort. Additionally, even though our cohort is one of the largest prospective bladder cancer cohorts, some SNPs had a very small MAF, and some important hits could have been lost. It is worth mentioning that we investigated the statistical interaction between genetic variants, smoking behavior, and the risk of bladder cancer. However, it does not directly imply a mechanistic interaction or causality. Our results highlight potential associations that may indicate the presence of complex biological interactions or shared pathways, but additional mechanistic studies are required to establish causality and plausible pathways.

Knowing these limitations, our results need to be investigated further in case-control or cohort studies; a GWAS meta-analysis would also be of utility to identify rare variants.

## 5. Conclusions

We found new suggestive loci on chromosomes 4q22.1, 12p13.1, and 16p13.3 interacting with different smoking phenotypes. Despite using the COBLAnCE cohort, one of the largest prospective bladder cancer cohorts, some limitations should be mentioned: use of a case-only design approach, absence of replication, and loss of information on some SNPs due to small MAF. Hence, these results need to be investigated in larger cohorts/case-control studies or within a meta-analysis framework.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cancers15174218/s1>, Figure S1: Pair-wise LD between SNPs in association with smoking phenotypes from the literature (shown in black) and the top SNPs from the COBLAnCE cohort (shown in blue). Red indicates a value of 1, and white indicates a value of 0. COBLAnCE, COhort to study BLadder CancEr. LD, linkage disequilibrium. SNP, single nucleotide polymorphisms; Table S1: Selected SNPs from the literature and their corresponding ORs and  $\beta$ s in the COBLAnCE cohort.

**Author Contributions:** M.K. and S.M.-P. performed the statistical analyses. M.K., S.B. and S.M. conceived and coordinated the study. K.G. provided overall study management. H.B., A.B., C.B. and J.-F.D. performed blood sample processing and genotyping. M.K. drafted the manuscript. All authors revised the manuscript for important intellectual content and approved the submission of the manuscript. M.K. had full access to the data and takes responsibility for the integrity of the data and the accuracy of the data analysis. The work reported in the paper has been performed by the authors unless clearly specified in the text. All authors have read and agreed to the published version of the manuscript.

**Funding:** This project is funded by the National Agency of Research (Programme Investissements d’Avenir Cohortes ANR-10-COHO-0007), the National Institute of Health and Medical Research (Soutien aux cohortes pour la Recherche sur le Cancer, AAP 2013), the National League against Cancer (EP 2011) and the National Cancer Institute (Research programme and interventions to reduce and fight against tobacco consumption 2019-1-TABAC). Sponsors were not included in the present study design, collection, analysis, and interpretation of data or in publications that resulted from the study.

**Institutional Review Board Statement:** The study protocol has been approved by the French regulatory authorities (data protection authority Commission Nationale de l’Informatique et des Libertés (CNIL) N° DR-2012-441, biological collections authority N° DC-2011-1486) and a French ethics committee (Comité de Protection des Personnes Ile de France VII (N° CO-12-001). This project was approved by one French ethics committee (Comité de Protection des Personnes Ile de France VII, reference number CO-12-001) for all participating centers: Hôpital Foch (Suresnes), Centre Hospitalier Universitaire Henri Mondor (Créteil), Centre Hospitalier Universitaire Gabriel-Montpied (Clermont-Ferrand), Centre Hospitalier Régional Universitaire de Lille, Centre Hospitalier Universitaire de Nîmes, Clinique Beau Soleil (Montpellier), Hopital Pasteur (Colmar), Centre Hospitalier René Dubos (Pontoise), Clinique du Parc Rambot (Aix en Provence), Centre Hospitalier Universitaire Grenoble Alpes (Grenoble), Clinique Urologique Atlantis (Nantes), Centre Hospitalier Européen Georges Pompidou (Paris), and Centre Hospitalier de Brabois (Nancy).

**Informed Consent Statement:** All patients were informed of the objectives of the study and were invited to voluntarily participate at inclusion. Patients who agreed to participate provided written consent.

**Data Availability Statement:** The GWAS summary statistics will be available in the NHGRI-EBI GWAS Catalog. Pseudoanonymised data and biological samples collected during the study can be obtained for conducting scientific projects. Requests should be made to the corresponding author and will be evaluated by the Scientific Committee of the COBLAnCE cohort. For approved projects, data transfer agreements will be required to be set up.

**Acknowledgments:** The authors would like to thank the following hospitals for their involvement in patient recruitment: Hôpital Foch (Suresnes), Centre Hospitalier Universitaire Henri Mondor (Créteil), Centre Hospitalier Universitaire Gabriel-Montpied (Clermont-Ferrand), Centre Hospitalier Régional Universitaire de Lille, Centre Hospitalier Universitaire de Nîmes, Clinique Beau Soleil (Montpellier), Hopital Pasteur (Colmar), Centre Hospitalier René Dubos (Pontoise), Clinique du Parc Rambot (Aix en Provence), Centre Hospitalier Universitaire Grenoble Alpes (Grenoble), Clinique Urologique Atlantis (Nantes), Centre Hospitalier Européen Georges Pompidou (Paris), and Centre Hospitalier de Brabois (Nancy), and Caroline Barau, Pascale Maille, Fannie Semprez and Pascale Soyeux at the Platform of Biological Resources and Research Team INSERM U955 of Henri Mondor Hospital, Elodie Guyon and Odette Mariani at the Biological Resources Centre of Curie Institute, Clémentine Krucker and Isabelle Bernard-Pierrot at UMR144 CNRS for the management of biological specimens, Marie-Claude Babron, Pascale Jan, Anthony Mangin and Pierre Guyader at Bureau de Biostatistique et d’Épidémiologie, Gustave Roussy for data management of the COBLAnCE study, Yves Koudou at

INSERM U1018 Oncostat for the quality control of genotyped data, and CEPH-Biobank and CNRGG genotyping teams for excellent technical support.

**Conflicts of Interest:** Julia Bonastre reports fees outside the scope of the submitted work as advisory board for BMS and funding of the MICADO study. Yves Allory reports fees outside the scope of the submitted work as consultant for AstraZeneca, Janssen, MSD, VitaDx, as invited speaker for Astellas, AstraZeneca, Bristol-Myers Squibb, MSD, Sanofi, as COBLAnCE cohort support for Bristol-Myers Squibb, MSD, and public and charity fundings from INCa, ANR, Ligue Contre le Cancer, and Fondation ARC. Simone Benhamou reports fundings outside the scope of the submitted work as COBLAnCE cohort support for Bristol-Myers Squibb, MSD, and public and charity fundings from INCa, ANR, and Ligue Contre le Cancer. Stefan Michiels reports fees outside the scope of the submitted work: Scientific Committee Study member: Roche, Data and safety monitoring member of clinical trials: Sensorion, Biophytis, Servier, IQVIA, Yuhan, Kedrion. The other authors declare no conflict of interest.

## References

- Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)] [[PubMed](#)]
- Yu, M.C.; Skipper, P.L.; Tannenbaum, S.R.; Chan, K.K.; Ross, R.K. Arylamine exposures and bladder cancer risk. *Mutat. Res. Mol. Mech. Mutagen.* **2002**, *506–507*, 21–28. [[CrossRef](#)] [[PubMed](#)]
- Chen, H.-I.; Liou, S.-H.; Loh, C.-H.; Uang, S.-N.; Yu, Y.-C.; Shih, T.-S. Bladder cancer screening and monitoring of 4,4'-Methylenebis(2-chloroaniline) exposure among workers in Taiwan. *Urology* **2005**, *66*, 305–310. [[CrossRef](#)] [[PubMed](#)]
- García-Pérez, J.; Pollán, M.; Boldo, E.; Pérez-Gómez, B.; Aragonés, N.; Lope, V.; Ramis, R.; Vidal, E.; López-Abente, G. Mortality due to lung, laryngeal and bladder cancer in towns lying in the vicinity of combustion installations. *Sci. Total. Environ.* **2009**, *407*, 2593–2602. [[CrossRef](#)] [[PubMed](#)]
- Fernández, M.I.; López, J.F.; Vivaldi, B.; Coz, F. Long-Term Impact of Arsenic in Drinking Water on Bladder Cancer Health Care and Mortality Rates 20 Years After End of Exposure. *J. Urol.* **2012**, *187*, 856–861. [[CrossRef](#)] [[PubMed](#)]
- Freedman, N.D.; Silverman, D.T.; Hollenbeck, A.R.; Schatzkin, A.; Abnet, C.C. Association Between Smoking and Risk of Bladder Cancer Among Men and Women. *JAMA* **2011**, *306*, 737–745. [[CrossRef](#)] [[PubMed](#)]
- Lammers, R.J.; Witjes, W.P.; Hendricksen, K.; Caris, C.T.; Janzing-Pastors, M.H.; Witjes, J.A. Smoking Status Is a Risk Factor for Recurrence After Transurethral Resection of Non-Muscle-Invasive Bladder Cancer. *Eur. Urol.* **2011**, *60*, 713–720. [[CrossRef](#)] [[PubMed](#)]
- Boström, P.J.; Alkhateeb, S.; Trottier, G.; Athanasopoulos, P.Z.; Mirtti, T.; Kortekangas, H.; Laato, M.; van Rhijn, B.; van der Kwast, T.; Fleshner, N.E.; et al. Sex differences in bladder cancer outcomes among smokers with advanced bladder cancer. *BJU Int.* **2011**, *109*, 70–76. [[CrossRef](#)]
- Saginala, K.; Barsouk, A.; Aluru, J.S.; Rawla, P.; Padala, S.A.; Barsouk, A. Epidemiology of Bladder Cancer. *Med Sci.* **2020**, *8*, 15. [[CrossRef](#)]
- Burger, M.; Catto, J.W.F.; Dalbagni, G.; Grossman, H.B.; Herr, H.; Karakiewicz, P.; Kassouf, W.; Kiemenev, L.A.; La Vecchia, C.; Shariat, S.; et al. Epidemiology and Risk Factors of Urothelial Bladder Cancer. *Eur. Urol.* **2013**, *63*, 234–241. [[CrossRef](#)]
- de Maturana, E.L.; Rava, M.; Anumudu, C.; Sáez, O.; Alonso, D.; Malats, N. Bladder Cancer Genetic Susceptibility. A Systematic Review. *Bl. Cancer* **2018**, *4*, 215–226. [[CrossRef](#)] [[PubMed](#)]
- Guey, L.T.; García-Closas, M.; Murta-Nascimento, C.; Lloreta, J.; Palencia, L.; Kogevinas, M.; Rothman, N.; Vellalta, G.; Calle, M.L.; Marenne, G.; et al. Genetic Susceptibility to Distinct Bladder Cancer Subphenotypes. *Eur. Urol.* **2010**, *57*, 283–292. [[CrossRef](#)] [[PubMed](#)]
- A Kiemenev, L.; Sulem, P.; Besenbacher, S.; Vermeulen, S.H.; Sigurdsson, A.; Thorleifsson, G.; Gudbjartsson, D.F.; Stacey, S.N.; Gudmundsson, J.; Zanon, C.; et al. A sequence variant at 4p16.3 confers susceptibility to urinary bladder cancer. *Nat. Genet.* **2010**, *42*, 415–419. [[CrossRef](#)]
- A Kiemenev, L.; Thorlacius, S.; Sulem, P.; Geller, F.; Aben, K.K.H.; Stacey, S.N.; Gudmundsson, J.; Jakobsdottir, M.; Bergthorsson, J.T.; Sigurdsson, A.; et al. Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat. Genet.* **2008**, *40*, 1307–1312. [[CrossRef](#)]
- Figuerola, J.D.; Ye, Y.; Siddiq, A.; Garcia-Closas, M.; Chatterjee, N.; Prokunina-Olsson, L.; Cortessis, V.K.; Kooperberg, C.; Cussenot, O.; Benhamou, S.; et al. Genome-wide association study identifies multiple loci associated with bladder cancer risk. *Hum. Mol. Genet.* **2013**, *23*, 1387–1398. [[CrossRef](#)] [[PubMed](#)]
- Rothman, N.; Garcia-Closas, M.; Chatterjee, N.; Malats, N.; Wu, X.F.; Figuerola, J.D.; Real, F.X.; Van den Berg, D.; Matullo, G.; Baris, D.; et al. A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. *Nat. Genet.* **2010**, *42*, 978–984. [[CrossRef](#)]
- Hein, D.W. Molecular genetics and function of NAT1 and NAT2: Role in aromatic amine metabolism and carcinogenesis. *Mutat. Res. Mol. Mech. Mutagen.* **2002**, *506–507*, 65–77. [[CrossRef](#)]

18. Figueroa, J.D.; Han, S.S.; Garcia-Closas, M.; Baris, D.; Jacobs, E.J.; Kogevinas, M.; Schwenn, M.; Malats, N.; Johnson, A.; Purdue, M.P.; et al. Genome-wide interaction study of smoking and bladder cancer risk. *Carcinog.* **2014**, *35*, 1737–1744. [[CrossRef](#)]
19. Koutros, S.; Kiemeny, L.A.; Choudhury, P.P.; Milne, R.L.; de Maturana, E.L.; Ye, Y.; Joseph, V.; Florez-Vargas, O.; Dyrskjot, L.; Figueroa, J.; et al. Genome-wide Association Study of Bladder Cancer Reveals New Biological and Translational Insights. *Eur. Urol.* **2023**, *84*, 127–137. [[CrossRef](#)]
20. Smith, P.G.; E Day, N. The Design of Case-Control Studies: The Influence of Confounding and Interaction Effects. *Leuk. Res.* **1984**, *13*, 356–365. [[CrossRef](#)]
21. Piegorsch, W.W.; Weinberg, C.R.; Taylor, J.A. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. *Stat. Med.* **1994**, *13*, 153–162. [[CrossRef](#)] [[PubMed](#)]
22. Gauderman, W.J. Sample Size Requirements for Association Studies of Gene-Gene Interaction. *Am. J. Epidemiology* **2002**, *155*, 478–484. [[CrossRef](#)] [[PubMed](#)]
23. Garcia-Closas, M.; Malats, N.; Silverman, D.; Dosemeci, M.; Kogevinas, M.; Hein, D.W.; Tardón, A.; Serra, C.; Carrato, A.; Garcia-Closas, R.; et al. NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: Results from the Spanish Bladder Cancer Study and meta-analyses. *Lancet* **2005**, *366*, 649–659. [[CrossRef](#)]
24. Garcia-Closas, M.; Rothman, N.; Figueroa, J.D.; Prokunina-Olsson, L.; Han, S.S.; Baris, D.; Jacobs, E.J.; Malats, N.; De Vivo, I.; Albanes, D.; et al. Common Genetic Polymorphisms Modify the Effect of Smoking on Absolute Risk of Bladder Cancer. *Cancer Res* **2013**, *73*, 2211–2220. [[CrossRef](#)] [[PubMed](#)]
25. Lubin, J.H.; Kogevinas, M.; Silverman, D.; Malats, N.; Garcia-Closas, M.; Tardón, A.; Hein, D.W.; Garcia-Closas, R.; Serra, C.; Dosemeci, M.; et al. Evidence for an intensity-dependent interaction of NAT2 acetylation genotype and cigarette smoking in the Spanish Bladder Cancer Study. *Leuk. Res.* **2007**, *36*, 236–241. [[CrossRef](#)]
26. Le Bret, T.; Bonastre, J.; Fraslin, A.; Neuzillet, Y.; Droupy, S.; Rebillard, X.; De la Taille, A.; Guy, L.; Villers, A.; Schneider, M.; et al. COBLAnCE: A Prospective Cohort to Study Prognostic and Predictive Factors in Bladder Cancer and to Generate Re-al-World Data on Treatment Patterns, Resource Use and Quality of Life. *BMJ Open* **2023**, *in press*.
27. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. *Tobacco Smoke and Involuntary Smoking*; IARC: Lyon, France, 2004.
28. Malhotra, J.; Borron, C.; Freedman, N.D.; Abnet, C.C.; Brandt, P.A.v.D.; White, E.; Milne, R.L.; Giles, G.G.; Boffetta, P. Association between Cigar or Pipe Smoking and Cancer Risk in Men: A Pooled Analysis of Five Cohort Studies. *Cancer Prev. Res.* **2017**, *10*, 704–709. [[CrossRef](#)]
29. Purcell, S.; Chang, C. PLINK, [v1.9]. Available online: [www.cog-genomics.org/plink/1.9/](http://www.cog-genomics.org/plink/1.9/) (accessed on 6 July 2023).
30. Chang, C.C.; Chow, C.C.; Tellier, L.C.; Vattikuti, S.; Purcell, S.M.; Lee, J.J. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* **2015**, *4*, 7. [[CrossRef](#)]
31. 1000 Genomes Project Consortium; Auton, A.; Brooks, L.D.; Durbin, R.M.; Garrison, E.P.; Kang, H.M.; Korbel, J.O.; Marchini, J.L.; McCarthy, S.; McVean, G.A.; et al. A global reference for human genetic variation. *Nature* **2015**, *526*, 68–74.
32. Delaneau, O.; McVean, G.A.; Donnelly, P.; Lunter, G.; Marchini, J.L.; Myers, S.; Gupta-Hinch, A.; Iqbal, Z.; Mathieson, I.; Rimmer, A.; et al. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat. Commun.* **2014**, *5*, 1–9. [[CrossRef](#)]
33. Howie, B.N.; Donnelly, P.; Marchini, J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLOS Genet.* **2009**, *5*, e1000529. [[CrossRef](#)] [[PubMed](#)]
34. Howie, B.; Marchini, J.; Stephens, M. Genotype Imputation with Thousands of Genomes. *G3 Genes | Genomes | Genetics* **2011**, *1*, 457–470. [[CrossRef](#)] [[PubMed](#)]
35. Wang, K.; Li, M.; Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **2010**, *38*, e164. [[CrossRef](#)] [[PubMed](#)]
36. Murcray, C.E.; Lewinger, J.P.; Gauderman, W.J. Gene-Environment Interaction in Genome-Wide Association Studies. *Am. J. Epidemiology* **2008**, *169*, 219–226. [[CrossRef](#)] [[PubMed](#)]
37. Liu, M.; Jiang, Y.; Wedow, R.; Li, Y.; Brazel, D.M.; Chen, F.; Datta, G.; Davila-Velderrain, J.; McGuire, D.; Tian, C.; et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat. Genet.* **2019**, *51*, 237–244. [[CrossRef](#)]
38. Sakornsakolpat, P.; Morrow, J.D.; Castaldi, P.J.; Hersh, C.P.; Bossé, Y.; Silverman, E.K.; Manichaikul, A.; Cho, M.H. Integrative genomics identifies new genes associated with severe COPD and emphysema. *Respir. Res.* **2018**, *19*, 46. [[CrossRef](#)]
39. Cheron, J.; D’exaerde, A.d.K. Drug addiction: From bench to bedside. *Transl. Psychiatry* **2021**, *11*, 1–14. [[CrossRef](#)]
40. Lin, J.; Peng, J.; Liu, G.; Deng, L. Overexpression of MECP2 attenuates cigarette smoke extracts induced lung epithelial cell injury by promoting CYP1B1 methylation. *J. Toxicol. Sci.* **2020**, *45*, 177–186. [[CrossRef](#)]
41. Korytina, G.F.; Akhmadishina, L.Z.; Kochetova, O.V.; Aznabaeva, Y.G.; Zagidullin, S.Z.; Victorova, T.V. Polymorphic variants of glutamate receptor (GRIK5, GRIN2B) and serotonin receptor (HTR2A) genes are associated with chronic. *Mol. Biol.* **2017**, *51*, 533–542. [[CrossRef](#)]
42. Vink, J.M.; Smit, A.B.; de Geus, E.J.; Sullivan, P.; Willemsen, G.; Hottenga, J.-J.; Smit, J.H.; Hoogendijk, W.J.; Zitman, F.G.; Peltonen, L.; et al. Genome-wide Association Study of Smoking Initiation and Current Smoking. *Am. J. Hum. Genet.* **2009**, *84*, 367–379. [[CrossRef](#)]



43. Pros, E.; Saigi, M.; Alameda, D.; Gomez-Mariano, G.; Martinez-Delgado, B.; Alburquerque-Bejar, J.; Carretero, J.; Tonda, R.; Esteve-Codina, A.; Catala, I.; et al. Genome-wide profiling of non-smoking-related lung cancer cells reveals common RB1 rearrangements associated with histopathologic transformation in EGFR-mutant tumors. *Ann. Oncol.* **2020**, *31*, 274–282. [[CrossRef](#)] [[PubMed](#)]
44. Wang, J.; Li, M.D. Common and Unique Biological Pathways Associated with Smoking Initiation/Progression, Nicotine Dependence, and Smoking Cessation. *Neuropsychopharmacology* **2009**, *35*, 702–719. [[CrossRef](#)] [[PubMed](#)]
45. Uhl, G.R.; Liu, Q.-R.; Drgon, T.; Johnson, C.; Walther, D.; Rose, J.E.; David, S.P.; Niaura, R.; Lerman, C. Molecular Genetics of Successful Smoking Cessation. *Arch. Gen. Psychiatry* **2008**, *65*, 683–693. [[CrossRef](#)]
46. Zhong, X.; Drgonova, J.; Li, C.-Y.; Uhl, G.R. Human cell adhesion molecules: Annotated functional subtypes and overrepresentation of addiction-associated genes. *Ann. New York Acad. Sci.* **2015**, *1349*, 83–95. [[CrossRef](#)] [[PubMed](#)]
47. Thorgeirsson, T.E.; Gudbjartsson, D.F.; Surakka, I.; Vink, J.M.; Amin, N.; Geller, F.; Sulem, P.; Rafnar, T.; Esko, T.; Walter, S.; et al. Sequence variants at CHRN3–CHRNA6 and CYP2A6 affect smoking behavior. *Nat. Genet.* **2010**, *42*, 448–453. [[CrossRef](#)]
48. Fernández-Castillo, N.; Gan, G.; van Donkelaar, M.M.; Vaht, M.; Weber, H.; Retz, W.; Meyer-Lindenberg, A.; Franke, B.; Harro, J.; Reif, A.; et al. RBFOX1, encoding a splicing regulator, is a candidate gene for aggressive behavior. *Eur. Neuropsychopharmacol.* **2017**, *30*, 44–55. [[CrossRef](#)]
49. O’leary, A.; Fernández-Castillo, N.; Gan, G.; Yang, Y.; Yotova, A.Y.; Kranz, T.M.; Grünewald, L.; Freudenberg, F.; Antón-Galindo, E.; Cabana-Domínguez, J.; et al. Behavioural and functional evidence revealing the role of RBFOX1 variation in multiple psychiatric disorders and traits. *Mol. Psychiatry* **2022**, *27*, 4464–4473. [[CrossRef](#)]
50. Bavarva, J.H.; Tae, H.; Settlage, R.E.; Garner, H.R. Characterizing the Genetic Basis for Nicotine Induced Cancer Development: A Transcriptome Sequencing Study. *PLoS ONE* **2013**, *8*, e67252. [[CrossRef](#)]
51. Samanic, C.; Kogevinas, M.; Dosemeci, M.; Malats, N.; Real, F.X.; Garcia-Closas, M.; Serra, C.; Carrato, A.; García-Closas, R.; Sala, M.; et al. Smoking and Bladder Cancer in Spain: Effects of Tobacco Type, Timing, Environmental Tobacco Smoke, and Gender. *Cancer Epidemiol. Biomark. Prev.* **2006**, *15*, 1348–1354. [[CrossRef](#)]
52. De Goeij, L.; Westhoff, E.; Witjes, J.A.; Aben, K.K.; Kampman, E.; Kiemeneij, L.A.; Vrieling, A. The UroLife study: Protocol for a Dutch prospective cohort on lifestyle habits in relation to non-muscle-invasive bladder cancer prognosis and health-related quality of life. *BMJ Open* **2019**, *9*, e030396. [[CrossRef](#)]
53. Zeegers, M.P.; Bryan, R.T.; Langford, C.; Billingham, L.; Murray, P.; Deshmukh, N.S.; Hussain, S.; James, N.; Wallace, D.M.A.; Cheng, K. The West Midlands Bladder Cancer Prognosis Programme: Rationale and design. *BJU Int.* **2010**, *105*, 784–788. [[CrossRef](#)] [[PubMed](#)]
54. Kwan, M.L.; Kushi, L.H.; Danforth, K.N.; Roh, J.M.; Ergas, I.J.; Lee, V.S.; Cannavale, K.L.; Harrison, T.N.; Contreras, R.; Loo, R.K.; et al. The Be-Well Study: A prospective cohort study of lifestyle and genetic factors to reduce the risk of recurrence and progression of non-muscle-invasive bladder cancer. *Cancer Causes Control.* **2019**, *30*, 187–193. [[CrossRef](#)] [[PubMed](#)]
55. Westhoff, E.; Wu, X.; Kiemeneij, L.A.; Lerner, S.P.; Ye, Y.; Huang, M.; Dinney, C.P.; Vrieling, A.; Tu, H. Dietary patterns and risk of recurrence and progression in non-muscle-invasive bladder cancer. *Int. J. Cancer* **2017**, *142*, 1797–1804. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.