Int J Mol Epidemiol Genet 2016;7(2):95-104
www.ijmeg.org /ISSN:1948-1756/IJMEG0027137

Original Article

A prospective and retrospective analysis of smoking behavior changes in ever smokers with high risk for lung cancer from New Mexico and Pennsylvania

Shuguang Leng[1], Joel L Weissfeld[2], Maria A Picchi[1], Mindi A Styn[3], Eric D Claus[4], Vincent P Clark[4], Guodong Wu[1], Cynthia L Thomas[1], Frank D Gilliland[5], Jianmin Yuan[2,6], Jill M Siegfried[7,8], Steven A Belinsky[1]

[1]Lung Cancer Program, Lovelace Respiratory Research Institute, Albuquerque, NM; [2]Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; [3]Thomas E. Starzl Transplantation Institute, University of Pittsburgh School of Medicine, Pittsburgh, PA; [4]Mind Research Network and Lovelace Biomedical and Environmental Research Institute, Albuquerque, NM; [5]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; [6]Division of Cancer Control and Population Sciences, University of Pittsburgh Cancer Institute, Pittsburgh, PA; [7]Department of Pharmacology & Chemical Biology, Hillman Cancer Center of The University of Pittsburgh Medical Center, Pittsburgh, PA; [8]Department of Pharmacology, University of Minnesota, Minneapolis, MN

Received March 2, 2016; Accepted May 1, 2016; Epub June 1, 2016; Published June 15, 2016

Abstract: Cigarette smoking is the leading preventable cause of death worldwide. The aim of this study is to conduct a prospective and retrospective analysis of smoking behavior changes in the Lovelace Smokers Cohort (LSC) and the Pittsburgh Lung Screening Study cohort (PLuSS). Area under the curve (AUC) for risk models predicting relapse based on demographic, smoking, and relevant clinical variables was 0.93 and 0.79 in LSC and PLuSS, respectively. The models for making a quit attempt had limited prediction ability in both cohorts (AUC≤0.62). We identified an ethnic disparity in adverse smoking behavior change that Hispanic smokers were less likely to make a quit attempt and were more likely to relapse after a quit attempt compared to non-Hispanic Whites. SNPs at 15q25 and 11p14 loci were associated with risk for smoking relapse in the LSC. Rs6495308 at 15q25 has a large difference in minor allele frequency between non-Hispanic Whites and Hispanics (0.46 versus 0.23, $P<0.0001$) and was associated with risk for ever relapse at same magnitude between the two ethnic groups (OR=1.36, 95% CI=1.10 to 1.67 versus 1.59, 95% CI=1.00 to 2.53, $P=0.81$). In summary, the risk prediction model established in LSC and PLuSS provided an excellent to outstanding distinguishing for abstainers who will or will not relapse. The ethnic disparity in adverse smoking behavior between Hispanics and non-Hispanic Whites may be at least partially explained by the sequence variants at 15q25 locus that contains multiple nicotine acetylcholine receptors.

Keywords: Smoking behavior, risk prediction model, ethnicity, single nucleotide polymorphism

Background

Cigarette smoking is the leading preventable cause of death worldwide, resulting in over 5 million deaths per year and an average loss of 15 years (y) of life in smokers [1]. Although smoking cessation benefits almost all smokers irrespective of the age at quitting or the cumulative amount of tobacco exposure, approximately 20% of adults in the United States continue to smoke [1]. Smoking cessation can be understood as a two-step process that involves making a quit attempt followed by maintaining abstinence. United States Food and Drug Administration has approved nicotine replacement therapy, bupropion, and varenicline as medication for smoking cessation that have shown modest pharmaceutical efficacy in addressing short term craving and withdrawal symptoms [1, 2]. In addition, smokers with reduced nicotine clearance capacity showed better response to transdermal nicotine therapy and varenicline treatment [2, 3]. However, the treatment efficacy and difference in response to treatment dissipated quickly once treatment stopped, and the overall long term

proportion (>6 month) for maintaining abstinence was still unacceptably low (≤15%) [2, 4, 1, 3]. Furthermore, extending treatment beyond the current treatment time (8-12 weeks) used in most smoking cessation trials is a concern because of the potential side effects from the cessation medications that include addiction and neuropsychiatric effects. Thus, the achievement of long-term abstinence by reducing relapses remains a major challenge for developing more effective smoking cessation strategies.

The development of risk prediction models for adverse smoking outcomes (e.g., continued smoking or smoking relapse) among ever smokers may improve smoking cessation outcomes by communicating a risk score and allocating available resources more efficiently. Results from studies using longitudinal smoker cohorts and lung cancer screening trials have found that younger age and shorter duration of smoking abstinence in former smokers were consistently associated with higher risk for smoking relapse [5-10]. In addition, the impact of a positive non-cancer screening outcome on smoking behavior appeared to be more prominent in promoting current smokers to make a quit attempt [6, 8, 10, 11]. A recent study defined smoking behavior change based on self-report in ever smokers enrolled in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) and developed a multivariate risk prediction model for smoking relapse in former smokers at study entry with an area under the curve (AUC) of a receiver operating characteristic (ROC) of 0.86 [7]. However, the smoking behavior changes were defined based on two questionnaires filled at study entry and one follow-up visit with a median interval of 8.5 years (range=4-14 years). The lack of repeated assessments of smoking behavior and the large time interval between the only two visits may compromise the ability to develop optimal prediction models and to understand the dynamics of smoking behavior changes.

Several genome-wide association studies (GWAS) have identified six loci [12-16] associated with quantitative measurements of nicotine dependence (8p11, 10q23, 15q25, and 19q13), smoking initiation (never versus ever smokers, 11p14), and smoking cessation (current versus former smokers, 9q34). However, the associations of these sequence variants with smoking behavior change carefully characterized in longitudinal cohorts that enroll moderate and heavy smokers have not been studied yet. Only one GWAS was conducted for smoking relapse and no loci reaching genome-wide significance were identified possibly due to the difficulty in differentiating smoking relapsers versus smokers who never made a successful quit attempt [17]. In this study, the smoking behavior changes were defined prospectively in the Lovelace Smokers Cohort (LSC) and the Pittsburgh Lung Screening Study (PLuSS) with repeated assessment of smoking behavior every 18 or 12 months, respectively. In addition, the smoking behavior changes were also defined retrospectively in the LSC due to the availability of the retrospective smoking data which qualitatively summarized the behavior changes in a person's smoking history. The associations of smoking behavior changes defined either prospectively or retrospectively with demographics, smoking history, clinical variables, and six known GWAS loci for nicotine addiction phenotypes were assessed using multivariate logistic regressions.

Methods

Study populations and study subjects

The LSC was established in 2001 to study biomarkers of chronic lung diseases including lung cancer in longitudinally collected biospecimens from smokers [18]. Enrollment was restricted to current and former smokers age 40 to 74 y with a minimum of 15 pack-years of smoking. Cohort members returned every 18 months to update the smoking and medical history. The current study included 2178 LSC members enrolled and followed through December 2012. The PLuSS Cohort was established in 2002 to support translational studies of the Pittsburgh Lung Cancer Specialized Programs of Research Excellence [11]. Eligibility criteria for inclusion were 50-79 years old; smoke half pack cigarettes per day or more for at least 25 years; if quit, smoking cessation was no more than 10 years; and no personal history of lung cancer. Enrollment of 3,638 persons was completed in 2005 and cohort members were contacted annually to update smoking and medical history (mainly cancer diagnosis). All study subjects signed a consent form and the Western Institutional Review Board and the Institutional Review Board for the University of Pittsburgh approved this project.

Definition for smoking relapse and quitting smoking in LSC

Smoking behavior change (smoking relapse and making a quit attempt) and time of event accurate to the month were collected in the follow-up questionnaire in LSC. Among baseline former smokers (n=740) with ≥1 follow-up visit, 51 subjects that reported resuming or restarting cigarette smoking in any follow-up visit were defined as prospective relapsers (PR). Non-PR (n=689) were defined as former smokers at baseline who did not report resuming cigarette smoking in any follow-up visit. We also included 33 additional PRs who quit smoking after enrollment and reported resuming smoking cigarettes in any succeeding follow-up visits. All former smokers reported smoking abstinence for ≥1 month. Among baseline current smokers (n=920) with ≥1 follow-up visit, 248 subjects that reported quitting cigarette smoking in any follow-up visit were defined as prospective quitters (PQ). Non-PQ (n=672) were defined as current smokers at baseline who did not report quitting cigarette smoking in any follow-up visit.

Retrospective relapsers (RR) and quitters (RQ) were defined based on answers to two questions asked at study entry in LSC: "Have you ever quit smoking for one year or longer?" (for current smokers only) and "Did you ever quit smoking for at least one year and then start smoking again?" (for former smokers only). Among 2178 LSC members, 1313 ever made a quit attempt were defined as RQs. The remaining participants (n=865) never made a quit attempt and were defined as non-RQs. Cohort members who were currently smoking at study entry and reported ever quitting for ≥1 y prior to study entry, were defined as RR (n=403). Former smokers who reported ever quitting smoking for ≥1 y prior to study entry and never resumed smoking were defined as non-RR (n=621). We also included 289 RRs who were former smokers at study entry but have taken ≥2 quit attempts before eventually quit smoking (answered "yes" to the second question). The phenotypes defined retrospectively qualitatively summarized the behavior changes in a person's smoking history. The robustness of the definitions was supported by the fact that only 4.8% non-RRs reported resuming smoking in any follow-up visit. Approximately 12.7% non-RQs reported making a quit attempt after enrollment and then maintained smoking abstinence ≥1 year.

Definition for PR and PQ in PLuSS

PR and PQ are defined based on the answer to a question in the PLuSS annual contact form: In the last 30 days, have you smoked any cigarettes? Among 1409 former smokers at study entry, 267 members who answered yes to this question in any follow-up contact were defined as PR. Among 2101 current smokers at study entry, 1224 members who answered no to this question were defined as PQ. Questions used for defining RR and RQ were not included in the baseline questionnaire in PLuSS.

Genotyping of the 15 GWAS single nucleotide polymorphisms (SNP) associated with nicotine addiction phenotypes

Genotype data for the 15 SNPs located within the six known GWAS loci for nicotine addiction phenotypes [12-16] were available for 1198 LSC subjects that contain 651 RRs and 363 non-RRs or 714 RQs and 422 non-RQs from our previous methylation GWAS [18]. Genotype data for additional LSC cohort members (346 RRs and 266 non-RRs) for three SNPs (rs4074134, rs6495308, and rs7103411) were acquired using TaqMan genotyping assays.

Statistical analysis

The associations between three categories of variables and risk for PR and odds ratio (OR) for PQ were assessed using logistic regression in LSC and PLuSS. These candidate factors in LSC included demographic variables (age, sex, ethnicity, and education), smoking related variables (time since quit, average cigarettes per day when smoking, age starting smoking, living with a smoker at home for ≥12 months during adult life, ever smoke less than usual amount for ≥12 months, and ever smoke more than usual amount for ≥12 months), and relevant clinical variables (body mass index, high blood pressure, heart trouble, diabetes, family history of COPD, family history of lung cancer, physician diagnosed emphysema or COPD, chronic bronchitis, and help available if needed during the past 4 weeks). Household income was not included in the model because of its correlation with education (spearman correlation coefficient=0.30, P<0.0001) and higher missing rate (21%). Duration of smoking, chronic lung disease, wheezing or whistling in the chest in the past 12 months, and baseline pulmonary function were not included in the model due to their

Table 1. Multivariate Logistic Regression Models of Smoking Relapse Defined Prospectively in LSC[a]

Variables	PR	Non-PR	OR	95% CI	*P* Value
n	84	682[b]			
Age (y, mean (SD))	54.4 (8.6)	61.1 (8.8)	0.74[c]	0.54, 1.02	0.065
Sex (Male, %)	11.9	22.8	0.40	0.17, 0.89	0.026
Education (per 1 level change of 6 levels, median (Q1, Q3))	4 (4, 5)	4 (4, 5)	1.28	0.96, 1.69	0.093
Hispanics (%)	11.9	10.0	1.85	0.72, 4.76	0.20
Time since quit (y, median (Q1, Q3))	1.0 (0.5, 1.5)	11.0 (4.6, 19.5)			
Spline 1					<0.0001
Spline 2					0.028
Spline 3					0.079
Ever smoking less cigarettes ≥1 y (%)	78.6	73.9	1.76	0.89, 3.48	0.11
Physician diagnosed emphysema or COPD (%)	6.0	15.8	0.26	0.091, 0.74	0.012
Baseline to relapse (y, median (Q1, Q3))	0.93 (0.36, 1.67)	Not applicable			

Abbreviations: CI, confidence interval; COPD, chronic obstructive pulmonary disease; LSC, Lovelace Smokers Cohort; OR, odds ratio; PR, prospective relapse; SD, standard deviation. [a]Area under ROC curve=0.929. [b]Six lacking education, one lacking physician diagnosed emphysema or COPD. [c]Odds ratio for every 10 years.

correlation with having physician diagnosed emphysema or COPD (spearman correlation coefficients>0.20, P<0.0001). Candidate factors in PLuSS included demographic variables (age, sex, ethnicity, education, marital status), smoking related variables (time since quit, average cigarettes per day when smoking, and smoking duration), and relevant clinical variables (body mass index, family history of lung cancer, previous personal cancer history, symptoms of hemoptysis, phlegm, cough, wheeze, dyspnea, edema, and weight loss, physician diagnosed bronchitis, emphysema, asthma, heart attack, stroke, coronary artery bypass or angioplasty, COPD, physician referral based on CT screening results, coronary calcification reported on screening CT, severity of airflow obstruction on study PFT, and time since most recent chest x-ray or CT before study entry). Using the most important CT finding, we classified subjects into four referral categories, including referral for moderate or high suspicion CT (greater than 5 percent predicted probability of lung cancer), referral for low suspicion CT (less than 5 percent predicted probability of lung cancer), referral for other reason (important CT finding not usually associated with lung cancer), and no referral. The first three referral groups were combined as one group in the analysis to maintain the statistical power.

Variables associated with risk for PR or odds ratio for PQ with P≤0.20 in univariate analysis were considered for inclusion in building the multivariate model. Stepwise selection with a significance level of 0.20 for allowing a variable to enter and stay in the model was used to create the most parsimonious model. Nonlinear effect of time since quit was evaluated with restricted cubic splines using four knots and three splines [19]. Knots were placed at 5, 35, 65, and 95 percentiles for time since quit in LSC and PLuSS to ensure adequate coverage of the entire data distribution [19, 20]. Model calibration was assessed by evaluating the deviation of the intercept and slope of the calibration line from the ideal values of 0 and 1, respectively when predicted probabilities were plotted vs observed probabilities.

Logistic regressions also assessed factors selected *a priori* for association with RR and RQ. Candidate factors included age and pack-years of smoking and those that occurred prior to relapse or quitting (sex, ethnicity, education, age starting smoking, respiratory illness during childhood, and living with a smoker for ≥12 months during adult life). The associations between 15 SNPs located within the six known GWAS loci for nicotine addiction phenotypes and risk for RR or odds ratio for RQ were analyzed using logistic regression with adjustment for covariates listed above. Each SNP was coded as 0, 1, and 2 for wild homozygote, heterozygote, and variant homozygote. All statistical tests were two-sided. Statistical analyses were conducted using SAS 9.2, R 2.14, and PLINK 1.06.

Results

Risk prediction model for smoking behavior change defined prospectively in LSC and PLuSS

The PR rate was 6.9% and 18.9% among former smokers at study entry in LSC and PLuSS,

Table 2. Multivariate Logistic Regression Models of Smoking Relapse Defined Prospectively in PLuSS[a]

Variables	PR	Non-PR	OR	95% CI	*P* Value
n	267	1142			
Age (y, mean (SD))	57.9 (6.2)	60.6 (6.9)	0.64[b]	0.50, 0.83	0.0005
Cigs/day			0.83	0.71, 0.98	0.026
1-19 Cigs/day	79 (29.6)	291 (25.5)			
20-29 Cigs/day	120 (44.9)	458 (40.1)			
30-39 Cigs/day	46 (17.2)	223 (19.5)			
40 or more Cigs/day	22 (8.2)	170 (14.9)			
Time since quit (y, median (Q1, Q3))	1 (0, 4)	5 (2, 8)			
Spline 1					<0.0001
Spline 2					0.039
Spline 3					0.24
Number of symptoms[c]			1.21	1.01, 1.44	0.036
None	76 (28.5)	408 (35.7)			
1	62 (23.2)	248 (21.7)			
2 or more	129 (48.3)	486 (42.6)			
Baseline COPD (yes, %)	76 (28.5)	394 (34.5)	0.78	0.55, 1.10	0.16

Abbreviations: CI, confidence interval; COPD, chronic obstructive pulmonary disease; OR, odds ratio; PLuSS, the Pittsburgh Lung Screening Study; PR, prospective relapse; SD, standard deviation. [a]Area under ROC curve=0.785. [b]Odds ratio for every 10 years. [c]Baseline symptoms, including phlegm, wheezing, shortness of breath, and ankle swelling.

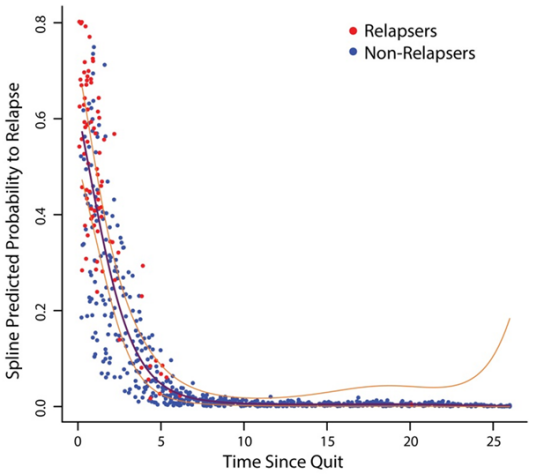

Figure 1. Nonlinear relationship between time since quit and the probability of relapse in LSC. The graph was prepared using restricted cubic splines with four knots at 0.4, 3.3, 18.5, 23.9 y. The *P* values for linear, quadratic, and cubic effects are <0.0001, 0.028, and 0.079, respectively. The purple line indicates the spline fit with 95% CIs indicated by the orange lines. Less than 10% members have time since quit >26 y, thus the time since quit data was truncated at 26 y to stabilize the variation in this figure.

respectively. Approximately, 74.5% and 68.9% relapse events occurred in cohort members who quit within 2.5 y prior to enrollment in LSC and PLuSS, respectively. The parsimonious logistic regression model contained multiple variables associated with risk for PR that overall provided an AUC ROC of 92.7% and 78.5% in LSC and PLuSS, respectively (**Tables 1** and **2**). The calibration line intercept and slope were 0 and 1 in both cohorts. The statistically significant predictors for PR were time since quit, history of physician diagnosed emphysema or COPD, sex, and age in LSC and time since quit, age, cigarettes smoked per day, and number of symptoms in PLuSS. The nonlinear relationship between time since quit and risk for PR was identified in both cohorts (**Tables 1** and **2**, **Figure 1**). The PR rate was significantly higher in cohort members who quit smoking within 2.5 y compared to those who quit greater than 2.5 y prior to study enrollment (27% versus 2.3% in LSC, and 38.4% versus 8.9% in PLuSS). Further analysis was restricted to former smokers who quit within 2.5 y, a population with higher risk for relapse. The covariates in the parsimonious logistic regression model provided a modest AUC ROC of 79.8% in LSC and 68.6% in PLuSS with time since quit as the most significant predictor (Supplementary Tables 1 and 2).

The PQ rate was 26.9% and 58.3% among baseline current smokers in LSC and PLuSS, respectively. The parsimonious logistic regression model for PQ provided an AUC ROC of

Table 3. Association between Six Known GWA Loci for Smoking Addition and Risk for Smoking Relapse and Quit Attempt in LSC

Gene name	Locus	SNP	Test SNP	R^2	Minor	MAF[a]	RR[b]			RQ[b]		
							OR	95% CI	*P* Value	OR	95% CI	*P* Value
CHRNA3, CHRNA5	15q25	rs1051730	Same	1	T	0.36	0.92	0.74, 1.15	0.46	1.14	0.94, 1.39	0.18
		rs16969968	Same	1	A	0.36	0.94	0.76, 1.18	0.61	1.15	0.95, 1.40	0.15
		rs55853698	rs2036527	1	T	0.37	0.90	0.72, 1.12	0.34	1.06	0.87, 1.29	0.55
		rs6495308	Same	1	C	0.22	1.35	1.12, 1.63	0.0017	0.93	0.80, 1.08	0.34
CHRNB3	8p11	rs6474412	rs4950	1	C	0.22	0.90	0.69, 1.17	0.44	1.01	0.81, 1.27	0.91
		rs13280604	rs4950	1								
Near PPP1R3C	10q23	rs1329650	Same	1	A	0.27	0.87	0.68, 1.11	0.27	0.86	0.70, 1.07	0.17
		rs1028936	Same	1	C	0.18	0.88	0.66, 1.18	0.40	0.86	0.67, 1.10	0.22
EGLN2, CYP2A6	19q13	rs3733829	Same	1	C	0.38	0.97	0.78, 1.22	0.81	0.92	0.76, 1.12	0.41
CYP2A6, CYP2B6		rs7937	Same	1	C	0.43	1.00	0.80, 1.25	0.99	1.03	0.85, 1.25	0.75
		rs1801272	NG									
		rs4105144	NG									
		rs7260329	Same	1	A	0.31	1.12	0.89, 1.42	0.33	1.01	0.83, 1.23	0.95
BDNF	11p14	rs6265	Same	1	A	0.18	0.80	0.61, 1.06	0.12	1.16	0.90, 1.49	0.24
		rs1013442	rs10734394	1	G	0.23	0.77	0.60, 1.00	0.052	1.17	0.93, 1.46	0.18
		rs4923457	rs4074134[c]	1	A	0.20	0.84	0.68, 1.03	0.092	0.95	0.81, 1.12	0.55
		rs4923460	rs4074134	1								
		rs4074134	Same	1								
		rs1304100	rs10734394	0.96								
		rs6484320	rs7103411[c]	0.87	C	0.20	0.83	0.68, 1.02	0.071	0.97	0.82, 1.14	0.71
		rs879048	rs4074134	1								
DBH	9q34	rs3025343	rs3025316	0.93	C	0.11	0.92	0.64, 1.31	0.63	0.80	0.60, 1.08	0.14

Abbreviations: CI, confidence interval; GWA, genome-wide association; MAF, minor allele frequency; OR, odds ratio; RQ, retrospective quit; RR, retrospective relapse; SNP, single nucleotide polymorphism. [a]MAF in 1163 NHWs. [b]Adjusted for age, sex, packyears, age start smoking, childhood Respiratory illness, Hispanic ethnicity, education, and not living with smokers during adulthood. Sample size is 651 RRs and 363 non-RRs and 714 RQs and 422 non-RQs for all 15 SNPs except for rs6495308, rs4074134, and rs7103411 that included additional 346 RRs and 266 non-RRs and 612 RQs and 451 non-RQs. RR: retrospective relapse. RQ: retrospective quit. [c]rs4074134 and rs7103411 are in high LD (r^2=0.91).

62.1% and 58.0% in LSC and PLuSS, respectively (Supplementary Tables 3 and 4). The calibration line intercept and slope were 0 and 1 in both cohorts. The statistically significant predictors for PQ were Hispanic ethnicity and sex in LSC and cigarettes smoked per day, marital status, and any medical conditions in PLuSS. Amount of time in cohort as a cohort related variable was also associated with odds ratio for making a quit attempt; odds ratio for PQ increased by 29.6% (95% confidence interval [CI]: 1.21, 1.39) and 20.5% (95% CI: 1.14, 1.28) for every 18 month interval in cohort in LSC and PLuSS, respectively.

Additional analyses were also conducted to assess whether previous smoking behavior change could affect the probability of smoking relapse or making a quit attempt after enrollment in LSC. The association between RR and PR was not statistically significant (OR=1.43, 95% CI: 0.78, 2.61). However, the association between RQ and PQ was highly statistically significant (OR=2.12, 95% CI: 1.55, 2.89).

Computerized tomography (CT) referral and smoking behavior change defined prospectively in PLuSS

Interestingly, the associations between the CT referral and risk for PR or odds ratio for PQ in PLuSS were not statistically significant ($P \geq 0.41$). Because the impact of physician referral due to abnormal screening outcomes on smoking behavior change appeared to be short-term, the analyses were repeated using the smoking status collected at the 1 year follow-up to redefine the PR and PQ. CT referral was associated with a 40% increased odds ratio for PQ (95% CI: 1.11, 1.76, not shown) with adjustment for the six variables listed in Supplementary Table 4. However, the association between CT referral and risk for PR remained statistically non-significant (OR=0.74, 95% CI: 0.46, 1.18, not shown). The findings further supported that CT referral in moderate and heavy smokers only had a short-term impact on promoting current smokers to make a quit attempt [6, 8, 10, 11].

Determinants for smoking behavior change defined retrospectively in LSC

The associations between eight variables and risk for RR and odds ratio for RQ in LSC are shown in Supplementary Tables 5 and 6, respectively. Interestingly, three variables including older age at enrollment, higher education level, and not living with smokers during adulthood were favorably associated with both measures toward quitting smoking. In addition, Hispanic smokers made fewer quit attempts and had difficulty maintaining smoking abstinence after quitting. Reanalysis of the association by considering the follow-up data in defining retrospective phenotypes did not change the results (not shown).

Association between SNPs and smoking behavior change defined retrospectively

Assessment of the 15 SNPs located within the six known GWAS loci for nicotine addiction phenotypes [12-16] in 651 RRs and 363 non-RRs or 714 RQs and 422 non-RQs (**Table 3**) identified significant associations between rs4074134 (OR=0.71, 95% CI: 0.54, 0.93, not shown), rs6495308 (OR=1.36, 95% CI: 1.05, 1.76, not shown), and rs7103411 (OR=0.71, 95% CI: 0.55, 0.93, not shown) and risk for RR. Genotyping additional cohort members (346 RRs and 266 non-RRs) for these three SNPs using TaqMan genotyping assay replicated the association between rs6495308 and risk for RR (OR=1.39, 95% CI: 1.05, 1.84, not shown). The pooled analysis resulted in a *P*-value of 0.0017 for rs6495308 that was below the significance level after Bonferroni correction (**Table 3**). Interestingly, rs6495308 had a large difference in MAF between Hispanics and non-Hispanic Whites (NHW, 0.46 versus 0.23, P< 0.0001, Supplementary Table 7) in LSC, though the magnitude of association between rs-6495308 and risk for RR did not differ by ethnicity (OR=1.36, 95% CI=1.10 to 1.67, P= 0.01 in NHWs, and odds ratio=1.59, 95% CI= 1.00 to 2.53, P=0.05 in Hispanics, P for interaction of rs6495308 and ethnicity=0.81).

Discussion

The prediction models that identify abstainers at risk for relapse and active smokers at greater chance for making a quit attempt were developed using two longitudinal cohorts that enrolled current and former smokers with high risk for lung cancer. With a comprehensive assessment of demographic, smoking related, and relevant clinical variables, the prediction model for smoking relapse developed had excellent to outstanding prediction accuracy. As the most significant determinant, longer time since quit was associated with reduced risk for smoking relapse in both cohorts. The nonlinear relationship between time since quit and risk for PR was replicated in both cohorts with 68.9-74.5% relapse events occurring in cohort members who quit within 2.5 y prior to enrollment. The relapse rate was 27-38.4% versus 2.3-8.9% in cohort members who quit within 2.5 y versus greater than 2.5 y prior to enrollment, respectively. The median interval between the baseline visit and relapse occurrence was only 0.93 y in LSC. Thus, these two longitudinal smoker cohorts provide a great source for future studies that will explore the mechanisms for smoking relapse in recent quitters at enrollment through the availability of detailed follow-up information for smoking behavior change and biospecimens collected at each visit. The prediction model for making a quit attempt developed had very minimal prediction ability (0.62 in LSC and 0.58 in PLuSS). The dramatically lower prediction performance in the model for making a quit attempt compared to the model for smoking relapse was also observed in the PLCO study [7].

The rates for smoking relapse among former smokers were 6.9 and 18.9% in LSC and PLuSS, respectively. The reported relapse rates across studies ranged from 3.3% to 10% [5-9]. The variation in relapse rates probably reflects the enrollment of former smokers with different average length of smoking abstinence prior to enrollment. PLuSS enrolled former smokers who quit cigarette smoking≤10 y prior to enrollment, while LSC has no restriction on this variable. The rate for quitting smoking among current smokers at study entry in LSC was 26.9% over an average of 5.3 y follow-up, a rate comparable to that (24%-35%) reported in long-term lung cancer screening trials [5, 10]. However, the quitting rate (58.3%) was quite high in PLuSS and this may be attributed to the older population with more comorbidity and long-term follow-up (9.4 y).

Among six known [12-16] loci associated with nicotine dependence, smoking initiation, and smoking cessation, 15q25 and 11p14 were

associated with risk for smoking relapse defined retrospectively in the LSC. Allele G for rs10734394, previously associated with reduced risk for being a regular smoker [15], was associated with reduced risk for smoking relapse (OR=0.77, *P*=0.052), suggesting that 11p14 may be a shared locus between smoking initiation and smoking relapse. Allele C of rs6495308, previously associated with lower cigarettes smoked per day [13], was associated with increased risk for smoking relapse (OR=1.35, *P*=0.0017), suggesting that lower nicotine addiction as assessed by cigarettes smoked per day may be a risk factor for smoking relapse. This premise seems contradictory to a popular notion that nicotine addiction is positively correlated with difficulty in smoking cessation [21]. However, both PLCO and PLuSS studies identified that greater cigarettes smoked per day, associated with higher risk for continuing smoking was also associated with lower risk for smoking relapse. Additional studies are needed to more precisely address the role of nicotine addiction mechanisms in smoking relapse. No SNPs within these six loci were associated with risk for smoking relapse or making a quit attempt defined prospectively in LSC (not shown), thus were not included in the prediction models.

The prospective and retrospective studies consistently showed that compared to NHWs, Hispanic smokers are less likely to make a quit attempt and are more likely to relapse after a quit attempt. This ethnic disparity in adverse smoking behavior is consistent with the observation that New Mexican Hispanics have higher risk for silencing of tumor suppressor genes in their lung and higher susceptibility for lung cancer risk [22]. The mechanism underlying this ethnic disparity, although largely unknown, could be attributed to culture and genetics. Stratification analysis by ethnicity in LSC identified significant associations between rs6495308 and risk for RR in NHWs and Hispanics, respectively. In addition, the magnitude of association between the two ethnic groups is not statistically significantly different (*P*=0.81). However, the allelic difference for rs6495308 between Hispanics and NHWs was highly statistically significant, suggesting the 15q25 locus should contribute to the genetic component responsible for the ethnic difference in smoking behavior change.

These results should be interpreted in the context of several limitations. First, smoking status in the longitudinal analysis was self-reported and not assessed using biochemical confirmation. However, research shows that self-report is a valid indicator of current smoking, especially when there are no strong incentives to deceive [23, 24]. Relative quantification of plasma nicotine and cotinine levels is available from a metabolomics study using 25 pairs of PRs and non-PRs in the LSC from this current study (Leng et al. unpublished data). A complete separation of pre- and post-relapse plasma samples was observed based on the nicotine and cotinine levels in these samples. Furthermore, approximately 95.5% agreement between self-reported abstinence status and CO measures (using 10 ppm as the cutoff for active smoking) was identified in a smoking cessation study that enrolled 161 smokers from New Mexico (Claus et al. unpublished data). These results strongly support self-report status used in this large scale epidemiological study as a sufficient indicator for current smoking. Second, information for cigar, pipe, and smokeless tobacco use (chewing tobacco and e-cigarette) was not collected until 2014 in LSC and was not included in the data analysis. However, we expect very minimal impact on the results because <3% LSC cohort members have ever reported using these nicotine containing products. Third, the PR and PQ were defined using the smoking status in the 30 days prior to completing the annual contact form in PLuSS. Thus, definition of non-PR and non-PQ were most vulnerable for error because smoking behavior changes between contacts were not collected. This may result in the reduced performance of the prediction models in PLuSS.

Conclusions

In summary, the risk prediction model for smoking relapse established in LSC and PLuSS provided an excellent external replication of the PLCO model with similar categories of variables. Second, the ethnic disparity in smoking behavior between Hispanics and NHWs may be at least partially explained by the sequence variants at 15q25 locus that contains multiple nicotine acetylcholine receptors.

Acknowledgements

This work was supported by a National Cancer Institute (NIH) R01 grant CA097356 and NIH/National Cancer Institute P30 grant CA118100.

The State of New Mexico as a direct appropriation from the Tobacco Settlement Fund provided initial support to establish the Lovelace Smokers Cohort. The PLuSS cohort was established and supported through the NCI SPORE in Lung Cancer grant P50 CA090440 to the University of Pittsburgh. We would like to thank Ms. Elise L. Calvillo at LRRI for the scientific editing of all figures.

Disclosure of conflict of interest

None.

Authors' contribution

SL and VC jointly designed the study. JW, MP, JY, JS, and SB led the fieldwork and collected the data. CT conducted DNA isolation and genotyping. SL, JW, MP, MS, and GW conducted the statistical analyses. SL drafted the manuscript. SL, JW, MP, MS, ED, FG, JY, and SB critical revised and all authors approved the manuscript.

Address correspondence to: Shuguang Leng, Lung Cancer Program, Lovelace Respiratory Research Institute, 2425 Ridgecrest Dr SE, Albuquerque, New Mexico 87108. Tel: 505-348-9160; Fax: 505-348-4990; E-mail: sleng@LRRI.org

References

[1] Rigotti NA. Smoking cessation in patients with respiratory disease: existing treatments and future directions. Lancet Respir Med 2013; 1: 241-50.

[2] Lerman C, Schnoll RA, Hawk LW Jr, Cinciripini P, George TP, Wileyto EP, Swan GE, Benowitz NL, Heitjan DF, Tyndale RF; PGRN-PNAT Research Group. Use of the nicotine metabolite ratio as a genetically informed biomarker of response to nicotine patch or varenicline for smoking cessation: a randomised, double-blind placebo-controlled trial. Lancet Respir Med 2015; 3: 131-8.

[3] Lerman C, Jepson C, Wileyto EP, Patterson F, Schnoll R, Mroziewicz M, Benowitz N, Tyndale RF. Genetic variation in nicotine metabolism predicts the efficacy of extended-duration transdermal nicotine therapy. Clin Pharmacol Ther 2010; 87: 553-7.

[4] Cahill K, Stevens S, Lancaster T. Pharmacological treatments for smoking cessation. JAMA 2014; 311: 193-4.

[5] Anderson CM, Yip R, Henschke CI, Yankelevitz DF, Ostroff JS, Burns DM. Smoking cessation and relapse during a lung cancer screening program. Cancer Epidemiol Biomarkers Prev 2009; 18: 3476-83.

[6] Ashraf H, Tonnesen P, Holst Pedersen J, Dirksen A, Thorsen H, Dossing M. Effect of CT screening on smoking habits at 1-year follow-up in the Danish Lung Cancer Screening Trial (DLCST). Thorax 2009; 64: 388-92.

[7] Barry SA, Tammemagi MC, Penek S, Kassan EC, Dorfman CS, Riley TL, Commin J, Taylor KL. Predictors of adverse smoking outcomes in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. J Natl Cancer Inst 2012; 104: 1647-59.

[8] Cox LS, Clark MM, Jett JR, Patten CA, Schroeder DR, Nirelli LM, Swensen SJ, Hurt RD. Change in smoking status after spiral chest computed tomography scan screening. Cancer 2003; 98: 2495-501.

[9] Gilpin EA, Pierce JP, Farkas AJ. Duration of smoking abstinence and success in quitting. J Natl Cancer Inst 1997; 89: 572-6.

[10] Townsend CO, Clark MM, Jett JR, Patten CA, Schroeder DR, Nirelli LM, Swensen SJ, Hurt RD. Relation between smoking cessation and receiving results from three annual spiral chest computed tomography scans for lung carcinoma screening. Cancer 2005; 103: 2154-62.

[11] Styn MA, Land SR, Perkins KA, Wilson DO, Romkes M, Weissfeld JL. Smoking behavior 1 year after computed tomography screening for lung cancer: Effect of physician referral for abnormal CT findings. Cancer Epidemiol Biomarkers Prev 2009; 18: 3484-9.

[12] Li MD, Burmeister M. New insights into the genetics of addiction. Nat Rev Genet 2009; 10: 225-31.

[13] Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton L, Berrettini W, Knouff CW, Yuan X, Waeber G, Vollenweider P, Preisig M, Wareham NJ, Zhao JH, Loos RJ, Barroso I, Khaw KT, Grundy S, Barter P, Mahley R, Kesaniemi A, McPherson R, Vincent JB, Strauss J, Kennedy JL, Farmer A, McGuffin P, Day R, Matthews K, Bakke P, Gulsvik A, Lucae S, Ising M, Brueckl T, Horstmann S, Wichmann HE, Rawal R, Dahmen N, Lamina C, Polasek O, Zgaga L, Huffman J, Campbell S, Kooner J, Chambers JC, Burnett MS, Devaney JM, Pichard AD, Kent KM, Satler L, Lindsay JM, Waksman R, Epstein S, Wilson JF, Wild SH, Campbell H, Vitart V, Reilly MP, Li M, Qu L, Wilensky R, Matthai W, Hakonarson HH, Rader DJ, Franke A, Wittig M, Schäfer A, Uda M, Terracciano A, Xiao X, Busonero F, Scheet P, Schlessinger D, St Clair D, Rujescu D, Abecasis GR, Grabe HJ, Teumer A, Völzke H, Petersmann A, John U, Rudan I, Hayward C, Wright AF, Kolcic I, Wright BJ, Thompson JR, Balmforth AJ, Hall

AS, Samani NJ, Anderson CA, Ahmad T, Mathew CG, Parkes M, Satsangi J, Caulfield M, Munroe PB, Farrall M, Dominiczak A, Worthington J, Thomson W, Eyre S, Barton A; Wellcome Trust Case Control Consortium, Mooser V, Francks C, Marchini J. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. Nat Genet 2010; 42: 436-40.

[14] Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, Sulem P, Rafnar T, Esko T, Walter S, Gieger C, Rawal R, Mangino M, Prokopenko I, Mägi R, Keskitalo K, Gudjonsdottir IH, Gretarsdottir S, Stefansson H, Thompson JR, Aulchenko YS, Nelis M, Aben KK, den Heijer M, Dirksen A, Ashraf H, Soranzo N, Valdes AM, Steves C, Uitterlinden AG, Hofman A, Tönjes A, Kovacs P, Hottenga JJ, Willemsen G, Vogelzangs N, Döring A, Dahmen N, Nitz B, Pergadia ML, Saez B, De Diego V, Lezcano V, Garcia-Prats MD, Ripatti S, Perola M, Kettunen J, Hartikainen AL, Pouta A, Laitinen J, Isohanni M, Huei-Yi S, Allen M, Krestyaninova M, Hall AS, Jones GT, van Rij AM, Mueller T, Dieplinger B, Haltmayer M, Jonsson S, Matthiasson SE, Oskarsson H, Tyrfingsson T, Kiemeney LA, Mayordomo JI, Lindholt JS, Pedersen JH, Franklin WA, Wolf H, Montgomery GW, Heath AC, Martin NG, Madden PA, Giegling I, Rujescu D, Järvelin MR, Salomaa V, Stumvoll M, Spector TD, Wichmann HE, Metspalu A, Samani NJ, Penninx BW, Oostra BA, Boomsma DI, Tiemeier H, van Duijn CM, Kaprio J, Gulcher JR; ENGAGE Consortium, McCarthy MI, Peltonen L, Thorsteinsdottir U, Stefansson K. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. Nat Genet 2010; 42: 448-53.

[15] Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat Genet 2010; 42: 441-7.

[16] Ware JJ, van den Bree MB, Munafo MR. Association of the CHRNA5-A3-B4 gene cluster with heaviness of smoking: a meta-analysis. Nicotine Tob Res 2011; 13: 1167-75.

[17] Tozzi F, Teumer A, Munafò M, Rawal R, Kazeem G, Gerbaulet M, McArdle W, Chilcoat H, Döring A, Dahmen N, Mooser V, Nauck M, Ring SM, Rubio JP, Vollenweider P, Waeber G, John U, Völzke H, Homuth G, Freyberger HJ, Völker U, Davey-Smith G, Gieger C, Preisig M, Grabe HJ. A genomewide association study of smoking relapse in four European population-based samples. Psychiatr Genet 2013; 23: 143-52.

[18] Leng S, Liu Y, Weissfeld JL, Thomas CL, Han Y, Picchi MA, Edlund CK, Willink RP, Gaither Davis AL, Do KC, Nukui T, Zhang X, Burki EA, Van Den Berg D, Romkes M, Gauderman WJ, Crowell RE, Tesfaigzi Y, Stidley CA, Amos CI, Siegfried JM, Gilliland FD, Belinsky SA. 15q12 variants, sputum gene promoter hypermethylation, and lung cancer risk: a GWAS in smokers. J Natl Cancer Inst 2015; 107.

[19] Harrell FE. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. Springer series in statistics. New York: Springer; 2001.

[20] Steyerberg EW, SpringerLink (Online service). Clinical Prediction Models A Practical Approach to Development, Validation, and Updating. Statistics for Biology and Health, New York, NY: Springer-Verlag New York; 2009.

[21] United States. Public Health Service. Office of the Surgeon General. How tobacco smoke causes disease: the biology and behavioral basis for smoking-attributable disease: a report of the Surgeon General. Rockville, MD. Washington, DC: U.S. Dept. of Health and Human Services, Public Health Service For sale by the Supt. of Docs., U.S. G.P.O.; 2010.

[22] Leng S, Liu Y, Thomas CL, Gauderman WJ, Picchi MA, Bruse SE, Zhang X, Flores KG, Van Den Berg D, Stidley CA, Gilliland FD, Belinsky SA. Native American ancestry affects the risk for gene methylation in the lungs of Hispanic smokers from New Mexico. Am J Respir Crit Care Med 2013; 188: 1110-6.

[23] SRNT Subcommittee on Biochemical Verification. Biochemical verification of tobacco use and cessation. Nicotine Tob Res 2002; 4: 149-59.

[24] Studts JL, Ghate SR, Gill JL, Studts CR, Barnes CN, LaJoie AS, Andrykowski MA, LaRocca RV. Validity of self-reported smoking status among participants in a lung cancer screening trial. Cancer Epidemiol Biomarkers Prev 2006; 15: 1825-8.

Supplementary Table 1. Multivariate logistic regression models of prospective smoking relapse in former smokers who quit smoking within 2.5 y prior to study entry in LSC[a]

Variables	PR	Non-PR	OR	95% CI	*P* Value
n	70	99			
Age (y, mean (SD))	54.2 (8.8)	57.4 (9.8)	0.65[b]	0.44, 0.97	0.035
Sex (Male, %)	12.9	25.3	0.32	0.11, 0.89	0.030
Education (per 1 level change of 6 levels, median (Q1, Q3))	4 (4, 5)	4 (3, 5)	1.17	0.81, 1.70	0.40
Hispanics (%)	12.9	10.1	2.94	0.88, 9.89	0.081
Quit-to-baseline (y, median (Q1, Q3))	0.8 (0.4, 1.2)	1.2 (0.9, 1.8)			
Spline 1					0.021
Spline 2					0.158
Spline 3					0.183
Ever smoking less cigarettes ≥1 y (%)	77.1	71.7	2.04	0.89, 4.69	0.092
Physician diagnosed emphysema or COPD (%)	4.3	22.2	0.13	0.033, 0.53	0.0044

Abbreviations: CI, confidence interval; COPD, chronic obstructive pulmonary disease; LSC, Lovelace Smokers Cohort; OR, odds ratio; PR, prospective relapse; SD, standard deviation. [a]Area under ROC curve=0.798. [b]Odds ratio per 10 years.

Supplementary Table 2. Multivariate logistic regression models of prospective smoking relapse in former smokers who quit smoking within 2.5 y prior to study entry in PLuSS[a]

Variables	PR	Non-PR	OR	95% CI	*P* Value
n	184	294			
Age (y, mean (SD))	57.7 (6.2)	59.0 (6.3)	0.72[b]	0.51, 1.01	0.057
Cigs/day			0.85	0.69, 1.06	0.14
1-19 Cigs/day	58 (31.5)	81 (27.6)			
20-29 Cigs/day	79 (42.9)	119 (40.5)			
30-39 Cigs/day	31 (16.9)	54 (18.4)			
40 or more Cigs/day	16 (8.7)	40 (13.6)			
Time since quit (y, median (Q1, Q3))[c]	0.5 (0, 1)	1 (0, 2)	0.48	0.37, 0.62	<0.0001
Number of symptoms[d]			1.22	0.96, 1.54	0.098
None	50 (27.2)	100 (34.0)			
1	40 (21.7)	57 (19.4)			
2 or more	94 (51.1)	137 (46.6)			
Baseline COPD (yes, %)	55 (29.9)	109 (37.1)	0.72	0.46, 1.13	0.15

Abbreviations: CI, confidence interval; COPD, chronic obstructive pulmonary disease; OR, odds ratio; PLuSS, the Pittsburgh Lung Screening Study; PR, prospective relapse; SD, standard deviation. [a]Area under ROC curve=0.686. [b]Odds ratio per 10 years. [c]Nonlinear effect of TSQ on risk for PR was not assessed due to less than 3 unique knots available when analysis was restricted to former smokers who quit≤2.5 y at baseline in PLuSS. However, this does not affect the integrity of the results because only linear effect was observed in former smokers who quit≤2.5 y in LSC. [d]Baseline symptoms, including phlegm, wheezing, shortness of breath, and ankle swelling.

Supplementary Table 3. Multivariate Logistic Regression Models of Making a Quit Attempt Defined Prospectively in LSC[a]

Variables	PQ	Non-PQ	OR	95% CI	*P* Value
n	247[b]	666[b]			
Age (y, mean (SD))	55.0 (8.5)	53.8 (8.7)	1.13[c]	0.95, 1.34	0.16
Sex (Male, %)	19.8	28.1	0.67	0.47, 0.96	0.029
Education (per 1 level change of 6 levels, median (Q1, Q3))	4 (3-5)	4 (3-5)	1.10	0.96, 1.27	0.17
Hispanics (%)	15.8	25.5	0.56	0.38, 0.84	0.0049
Cigs/day (10 cigs, mean (SD))	19.2 (8.8)	20.7 (8.7)	0.85	0.71, 1.02	0.078
Ever smoking more cigarettes ≥1 y (%)	61.9	55.6	1.25	0.92, 1.71	0.15
CMH (%)	36.4	44.0	0.76	0.56, 1.04	0.086
Baseline to quit (y, median (Q1, Q3))	2.59 (1.18, 4.10)	Not applicable			

Abbreviations: CI, confidence interval; CMH, chronic mucous hypersecretion; LSC, Lovelace Smokers Cohort; OR, odds ratio; PQ, prospective quit; SD, standard deviation.
[a]Area under ROC curve=0.62. [b]One PQ and three non-PQs have missing education. Three non-PQs have missing CMH1. [c]Odds ratio for every 10 years.

Supplementary Table 4. Multivariate Logistic Regression Models of Making a Quit Attempt Defined Prospectively in PLuSS[a]

Variables	PQ	Non-PQ	OR	95% CI	*P* Value
n	1224	877			
Age (y, mean (SD))	58.8 (6.7)	58.4 (6.5)	1.13[b]	0.99, 1.29	0.077
Sex (male, %)	581 (47.5)	447 (51.0)	0.88	0.74, 1.06	0.19
Cigs/day			0.81	0.73, 0.89	<0.0001
1-19 Cigs/day	473 (38.6)	265 (30.2)			
20-29 Cigs/day	528 (43.1)	392 (44.7)			
30-39 Cigs/day	155 (12.7)	152 (17.3)			
40 or more Cigs/day	68 (5.6)	68 (7.8)			
Ethnicity (NHW, %)	1156 (94.4)	812 (92.6)	1.19	1.03, 1.37	0.063
Marital status			0.84	0.73, 0.97	0.016
Married or living as married	789 (64.5)	527 (60.1)			
Widowed, divorced, or separated	361 (29.5)	266 (30.3)			
Never married	74 (6.1)	84 (9.6)			
Any medical condition (yes, %)[c]	367 (30.0)	306 (34.9)	0.81	0.68, 0.99	0.040

Abbreviations: CI, confidence interval; NHW, non-Hispanic White; OR, odds ratio; PLuSS, the Pittsburgh Lung Screening Study; PQ, prospective quit; SD, standard deviation. [a]Area under ROC curve=0.580. [b]Odds ratio for every 10 years. [c]Baseline physician diagnoses, including chronic bronchitis, emphysema, and asthma.

Supplementary Table 5. Multivariate logistic regression models of ever having a smoking relapse defined retrospectively in LSC

Variables	RR	Non-RR	OR	95% CI	*P* Value
n	688[a]	612[a]			
Baseline age (y, mean (SD))	56.0 (9.2)	60.8 (9.1)	0.58[b]	0.51, 0.67	<0.0001
Sex (Male, %)	21.8	24.8	0.94	0.71, 1.25	0.69
Baseline packyears (median (Q1, Q3))	31.6 (24.0, 42.8)	34.3 (25.0, 50.2)	0.99[c]	0.96, 1.02	0.49
Age start smoking (y, mean (SD))	16.5 (3.5)	16.9 (4.1)	1.06[d]	0.90, 1.25	0.47
Childhood respiratory illness (%)	10.0	11.3	0.97	0.67, 1.40	0.86
Hispanics (%)	19.8	9.6	1.97	1.39, 2.79	0.00014
Education (per 1 level change of 6 levels, median (Q1, Q3))	4 (3, 5)	4 (4, 5)	0.89	0.80, 0.99	0.033
Not living with smokers during adulthood (%)	13.4	16.8	0.70	0.51, 0.97	0.033

Abbreviations: CI, confidence interval; LSC, Lovelace Smokers Cohort; OR, odds ratio; RR, retrospective relapse; SD, standard deviation. [a]Two RRs and one non-RR have missing Respiratory illness at childhood. Two RRs and six non-RRs have missing education. Two non-RRs have missing age. [b]Odds ratio per 10 years. [c]Odds ratio per five packyears. [d]Odds ratio per five years.

Supplementary Table 6. Multivariate logistic regression models of ever making quit attempt defined retrospectively in LSC

Variables	RQ	Non-RQ	OR	95% CI	*P* Value
n	1300[a]	861[a]			
Baseline age (y, mean (SD))	58.2 (9.5)	53.5 (8.7)	2.38[b]	2.10-2.70	<0.0001
Sex (Male, %)	23.2	25.4	0.89	0.71-1.11	0.31
Baseline packyears (median (Q1, Q3))	32.8 (24.5, 46.0)	37.5 (28.2, 50.0)	0.87[c]	0.85-0.90	<0.0001
Age start smoking (y, mean (SD))	16.7 (3.8)	16.7 (3.8)	0.63[d]	0.55-0.72	<0.0001
Childhood Respiratory illness (%)	10.6	8.8	1.13	0.82-1.56	0.44
Hispanics (%)	15.1	22.7	0.73	0.57-0.94	0.014
Education (per 1 level change of 6 levels, median (Q1, Q3))	4 (4, 5)	4 (3, 4)	1.43	1.31-1.57	<0.0001
Not living with smokers during adulthood (%)	15.1	10.3	1.47	1.10-1.97	0.0089

Abbreviations: CI, confidence interval; LSC, Lovelace Smokers Cohort; OR, odds ratio; RQ, retrospective quit; SD, standard deviation. [a]Three RQs have missing Respiratory illness at childhood. Eight RQs and four non-RQs have missing education. Two RQs have missing age. [b]Odds ratio per 10 years. [c]Odds ratio per five packyears. [d]Odds ratio per five years.

Supplementary Table 7. Ethnic difference for three SNPs in LSC

Gene name	Locus	SNP	Allele	MAF		
				NHWs (n=1640)	Hispanics (n=402)	*P* Value[a]
CHRNA3, CHRNA5	15q25	rs6495308	T/C	0.23	0.46	<0.0001
BDNF	11p14	rs4074134	G/A	0.20	0.23	0.097
BDNF	11p14	rs7103411	T/C	0.20	0.23	0.35

Abbreviations: LSC, Lovelace Smokers Cohort; MAF, minor allele frequency, NHW, non-Hispanic White; SNP, single nucleotide polymorphism. [a]Chi square test based on genotype distribution.