ORIGINAL ARTICLE

# Polygenic Risk and the Developmental Progression to Heavy, Persistent Smoking and Nicotine Dependence

## *Evidence From a 4-Decade Longitudinal Study*

*Daniel W. Belsky, PhD; Terrie E. Moffitt, PhD; Timothy B. Baker, PhD; Andrea K. Biddle, PhD; James P. Evans, MD, PhD; HonaLee Harrington, BA; Renate Houts, PhD; Madeline Meier, PhD; Karen Sugden, PhD; Benjamin Williams, BS; Richie Poulton, PhD; Avshalom Caspi, PhD*

**Importance:** Genome-wide hypothesis-free discovery methods have identified loci that are associated with heavy smoking in adulthood. Research is needed to understand developmental processes that link newly discovered genetic risks with adult heavy smoking.

**Objective:** To test how genetic risks discovered in genome-wide association studies of adult smoking influence the developmental progression of smoking behavior from initiation through conversion to daily smoking, progression to heavy smoking, nicotine dependence, and struggles with cessation.

**Design:** A 38-year, prospective, longitudinal study of a representative birth cohort.

**Setting:** The Dunedin Multidisciplinary Health and Development Study of New Zealand.

**Participants:** The study included 1037 male and female participants.

**Exposure:** We assessed genetic risk with a multilocus genetic risk score. The genetic risk score was composed of single-nucleotide polymorphisms identified in 3 meta-analyses of genome-wide association studies of smoking quantity phenotypes.

**Main Outcomes and Measures:** Smoking initiation, conversion to daily smoking, progression to heavy smoking, nicotine dependence (Fagerström Test of Nicotine Dependence), and cessation difficulties were evaluated at 8 assessments spanning the ages of 11 to 38 years.

**Results:** Genetic risk score was unrelated to smoking initiation. However, individuals at higher genetic risk were more likely to convert to daily smoking as teenagers, progressed more rapidly from smoking initiation to heavy smoking, persisted longer in smoking heavily, developed nicotine dependence more frequently, were more reliant on smoking to cope with stress, and were more likely to fail in their cessation attempts. Further analysis revealed that 2 adolescent developmental phenotypes—early conversion to daily smoking and rapid progression to heavy smoking—mediated associations between the genetic risk score and mature phenotypes of persistent heavy smoking, nicotine dependence, and cessation failure. The genetic risk score predicted smoking risk over and above family history.

**Conclusions and Relevance:** Initiatives that disrupt the developmental progression of smoking behavior among adolescents may mitigate genetic risks for developing adult smoking problems. Future genetic research may maximize discovery potential by focusing on smoking behavior soon after smoking initiation and by studying young smokers.

*JAMA Psychiatry. 2013;70(5):534-542.*
*Published online March 27, 2013.*
*doi:10.1001/jamapsychiatry.2013.736*

**Author Affiliations** are listed at the end of this article.

CIGARETTE SMOKING IS A costly, prevalent public health problem. The US Centers for Disease Control and Prevention attribute more than 400 000 deaths and $95 million in lost productivity to smoking during 2000-2004.[1] Approximately 20% of adults still smoke daily despite widespread knowledge of smoking's health effects and increasing economic costs to smokers due to increasing taxes.[2] Thus, more effective interventions to prevent smoking, motivate smoking cessation, and prevent relapse are needed.[3-5]

Studies of twins[6] suggest that genetic differences among individuals have an important role in smoking behavior, cessation, and response to antismoking interventions. Recent genome-wide association studies (GWASs)[7-9] in adult smokers and former smokers revealed genes that re-

late with genome-wide significance to smoking quantity (number of cigarettes smoked per day). These genes are already being used in clinical applications (eg, to predict smoking cessation likelihood and in pharmacogenetic analyses).[10-14] An important additional step in the translation of these GWAS findings is to test whether genetic markers that predicted smoking quantity in GWASs also predict the development of smoking behavior in adolescence.[15,16] This question is of critical importance for public health practice because intervention to disrupt genetic risk is likely to be most effective early in the development of dependence. Important developmental phenotypes in the pathogenesis of adult dependence include smoking initiation, conversion to daily smoking during adolescence, and rapid progression to heavy smoking.[17] Early, rapid progression from smoking initiation to heavy use is a signal risk for adult nicotine dependence.[18-21] Therefore, the present study tested relations of GWAS-identified genetic risk with adolescent and adult smoking phenotypes and then determined the extent to which genetic effects on the former affected the adult phenotype outcomes.

In this study, we tested prospective associations between genetic risks and adolescent developmental and mature adult phenotypes of smoking behavior (**Figure 1**). We examined genetic risks in the Dunedin Study, a birth cohort (n = 1037) followed up to the age of 38 years with 95% retention. We collected smoking behavior data at 8 assessments spanning the ages of 11 to 38 years. This approach allowed us to study the effects of genetic risk in the cohort as members initiated smoking during adolescence, converted to daily smoking, and progressed to heavy smoking during the teenage and young adult years and as they developed nicotine dependence and struggled with cessation in their 20s and 30s. We tested whether individuals at higher genetic risk progressed more rapidly from smoking initiation to heavy smoking, if they smoked more heavily as adults, if they were more nicotine dependent, and if they were more likely to fail in their cessation attempts. Finally, we tested the hypothesis that genetic risk accelerates the developmental progression from smoking initiation to heavy smoking, and this, in turn, increases the severity of adult smoking problems, such as heavy, intractable smoking and nicotine dependence. This model has relevance to public health interventions that might delay the developmental progression to heavy smoking. To put the magnitudes of genetic risk effects in context and to determine whether molecular genetic measurements provided novel information about risk, we conducted an additional analysis comparing molecular genetic information to family history information. These analyses asked how large molecular genetic effects were relative to family history effects and whether molecular genetic effects were independent of family history effects in predicting risk.

## METHODS

### SAMPLE

Participants are members of the Dunedin Multidisciplinary Health and Development Study, a longitudinal investigation of health and behavior in a complete birth cohort. Study members (N = 1037, 91% of eligible births, 52% male) were all individuals born between April 1972 and March 1973 in Dunedin, New Zealand, who were eligible for the longitudinal study based on residence in the province at age 3 years and who participated in the first follow-up assessment at age 3 years. The cohort represents the full range of socioeconomic status in the general population of New Zealand's South Island and is primarily white.[22] Assessments were performed at birth and at the ages of 3, 5, 7, 9, 11, 13, 15, 18, 21, 26, 32, and, most recently, 38 years, when 1007 study members were still alive, with 95% retention. At each assessment wave, study members are brought to the Dunedin research unit for a full day of interviews and examinations. The Otago Ethics Committee approved each phase of the study, and informed consent was obtained from all study members.

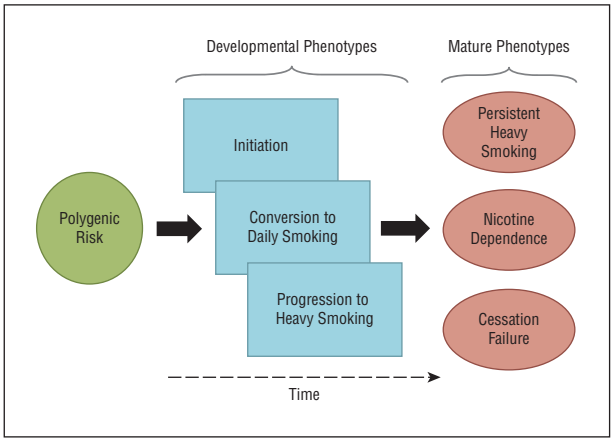



**Figure 1.** Genetic risk and the developmental progression of smoking behavior. In the hypothesized model, genetic risk influences the mature phenotypes of heavy smoking persistence, nicotine dependence, and cessation failure through a pathway mediated by 3 developmental phenotypes: smoking initiation, conversion to daily smoking, and progression to heavy smoking.

### MEASURES

#### Genetic Risk Score

A challenge for developmental research following up GWAS discoveries is that effect sizes for individual single-nucleotide polymorphisms (SNPs) are small; the largest effects for smoking quantity approach a change of 1 cigarette per day per risk allele. Moreover, many of the longitudinal studies[23] with data necessary to investigate developmental phenotypes are underpowered to test individual SNP effects. However, evidence shows that smoking-associated loci make additive contributions to risk, recommending aggregating risk alleles.[24-27] Summing risk alleles across GWAS-identified SNPs to compute a genetic risk score (GRS) yields a quantitative index of genetic risk with a normal distribution[28] and a potentially larger effect size.

We derived the GRS from 3 recent meta-analyses of GWAS that used as their phenotype cigarettes smoked per day.[7-9] To construct the GRS, we considered SNPs from regions with genome-wide significant associations in at least 2 meta-analyses. All 3 meta-analyses identified SNPs in the q25.1 region of chromosome 15 containing the *CHRNA5-CHRNA3-CHRNB4* gene cluster. Two meta-analyses identified SNPs in the q13.2 region of chromosome 19 containing the gene *CYP2A6*. These genes influence nicotine response and nicotine metabolism, have been linked with nicotine dependence, and are candidate genes in research into the development of smoking behavior.[26,29-35] Therefore, we focused our inquiry on the top GWAS SNPs in

these 2 regions (eMethods; http://www.jamapsych.com). In 15q25.1, we selected the SNPs rs16969968, rs6495308, rs8032771, and rs12595538. The SNPs rs16969968 and rs6495308, which fall within the *CHRNA5-CHRNA3-CHRNB4* gene cluster, were reported previously to have independent associations with smoking quantity.[8,36] The SNPs rs8032771 and rs12595538, which are located downstream of the *CHRNA5-CHRNA3-CHRNB4* gene cluster, were in weak linkage disequilibrium with rs16969968 and rs6495308 ($R^2 \leq 0.10$) and were genome-wide significant in the largest meta-analysis[7] ($P < 1 \times 10^{-16}$ for both; *P* values for these SNPs were not published in the other 2 meta-analyses). In 19q13.2, we selected the SNPs rs7937 and rs4105144. Following 2 previous studies[25,27] using multilocus measures of genetic risk for smoking, we assumed an additive model and summed alleles associated with higher smoking quantity to calculate the GRS. Because no reference data exist to determine the exact contributions of individual SNPs in our GRS to developmental phenotypes of smoking behavior, we used unweighted counts of risk alleles to construct the score.

To validate this GRS, we used independent data from the Atherosclerosis Risk in the Communities database and the Study of Addiction: Genetics and Environment database, accessed through the National Institutes of Health Database of Genotypes and Phenotypes.[37,38] When a GRS SNP was unavailable in one of these databases, we selected the closest linkage disequilibrium proxy for that SNP to include in the GRS. Among European-descent Atherosclerosis Risk in the Communities participants (n=8293), each SD increase in the GRS predicted a 1.45-pack-year increase in lifetime cigarette consumption among individuals who had ever smoked ($P < .001$) and a 1.02-cigarette increase in daily consumption among these ever smokers ($P < .001$). Replication of the GRS–smoking quantity association in the Study of Addiction: Genetics and Environment database and additional validation analyses testing versions of the GRS that exclude the SNPs rs16969968 and rs6495308 are presented in eTable 1.

Dunedin cohort genotyping was conducted with a commercially available array (BeadPlex Array; Illumina, Inc) using DNA extracted from whole blood (93% of the sample) or buccal swabs (7% of the sample). The GRS SNPs or proxies (linkage $R^2 \geq 0.85$) were called successfully in 95% of European-descent study members (eTable 2). These 880 individuals formed the analysis sample. Cohort members carried a mean (SD) of 7.06 (2.27) of 12 possible risk alleles. Cohort members' sex and socioeconomic status[39] were unrelated to their genetic risk (Pearson $r \leq 0.01$). The GRS was standardized to have a mean (SD) of 0 (1) for analyses (GRS).

### Family History of Smoking

Family histories of smoking were available for 99% of the cohort. The family history consisted of reports of smoking history provided by study members and both parents for study members' siblings, parents, and grandparents. The family history was summarized as the proportion of family members in the pedigree who were ever regular smokers, adjusted to account for differences in genetic relatedness to the proband of first- and second-degree relatives.[40]

### Smoking Behavior

The developmental progression of smoking behavior in the Dunedin cohort is shown in **Figure 2**A. Measurement of adolescent developmental phenotypes and mature phenotypes of smoking behavior is shown in Figure 2B.

### STATISTICAL ANALYSIS

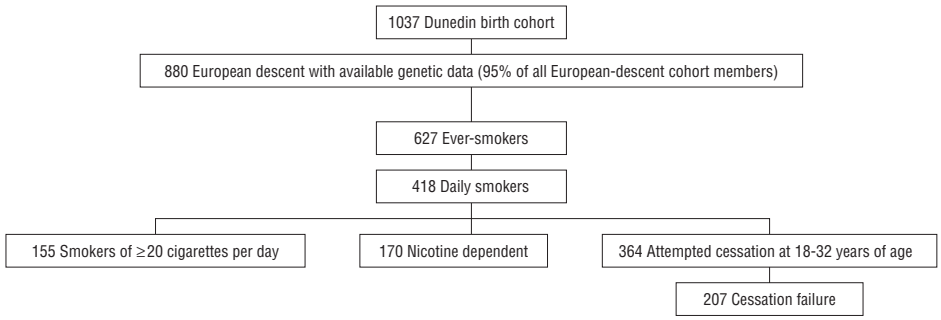
Data analysis was divided into 3 parts. First, we analyzed associations between the GRS and developmental phenotypes of smoking behavior. Second, we analyzed associations between the GRS and mature phenotypes. Third, we tested whether developmental phenotypes mediated associations between the GRS and mature phenotypes. We used different statistical models to analyze outcome data as required by the outcome's distribution. We analyzed continuously distributed outcome data (eg, lifetime cigarette consumption in pack-years) using ordinary least squares. We analyzed dichotomous outcome data (eg, daily smoker by age 15 years) using Poisson regression models because this is a standard method to derive relative risks.[47] We analyzed count outcome data (eg, the number of assessments at which the study member met criteria for nicotine dependence) using negative binomial regression models to account for the overdispersion of many of the count measures.[48] We analyzed hazards of smoking initiation, progression to heavy smoking, becoming nicotine dependent, and relapsing from a quit attempt using Cox proportional hazards regression models. To account for differences in the frequency with which study members attempted cessation, we constructed panel data sets that included one observation per study member per assessment (for the data for ages 18-32 years) and one observation per study member per quit attempt (for the data for life-history calendars). We used these panel data sets to analyze the genetic effect on smokers' risks of cessation failure during ages 18 to 32 years and on their hazards of relapse during ages 32 to 38 years. We accounted for nonindependence of repeated observations of individuals using generalized estimating equation models of risks and conditional risk-set models of hazards.[49,50] We tested whether genetic effects on the mature phenotypes of persistent heavy smoking, nicotine dependence, and relapse were mediated by adolescent developmental phenotypes using the structural equation described by McKinnon and Dwyer[51] and the methods described by Preacher et al.[52,53] To allow for a single test of mediation, we conducted a principal components analysis[54] of the mature phenotypes of persistent heavy smoking (pack-years smoked at age 38 years), nicotine dependence (total number of symptoms across all assessments), and cessation failure (number of assessments with relapse). This analysis indicated that the mature phenotypes were positively and significantly correlated (eTable 3) and could be summarized in a single component that explained 78% of the variance in the 3 measures (factor loading = 0.61 for persistent heavy smoking, 0.60 for nicotine dependence, and 0.52 for cessation failure). We used this component as the dependent variable in our mediation analysis. Analyses were adjusted for sex and conducted using STATA statistical software, version 11.0 (StataCorp LP).[55] Panel-data models were fitted to longitudinal repeated-measures data using the XT and ST commands in STATA statistical software, version 11.0. Unless otherwise noted, effect sizes are presented for 1-SD increase in genetic risk.

## RESULTS

### GENETIC RISK AND SMOKING INITIATION

The GRS was not associated with whether individuals initiated smoking or with the timing of initiation (relative risk [RR] for smoking initiation = 0.98; 95% CI, 0.95-1.02; cumulative hazard ratio [HR] for initiation = 1.01; 95% CI, 0.94-1.09; based on a 1-SD increase in genetic risk; **Table**). Subsequent analyses focused on the 627

A




B

| Developmental Phenotypes | No. | (%) |
|---|---|---|
| **Initiation.** Age at which study members reported first smoking at least occasionally. Survival time to initiation was calculated as the age at which a study member first began smoking at least occasionally. The 627 cohort members who initiated smoking represent 71% of the 880 European-descent cohort members with available genetic data. | 627 | |
| **Conversion Daily Smoking.** The first assessment at which a study member smoked ≥1 cigarette per day. | 418 | (67) |
| **Early Conversion to Daily Smoking.** Daily smoker at the 15 years of age assessment. Most study members who ever converted to daily smoking did so by the 18 years of age assessment (74.4%). Therefore, we defined "early conversion" to daily smoking as having converted by the previous assessment at 15 years of age. | 121 | (19) |
| **Progression to Smoking ≥20 Cigarettes per Day.** Ever a smoker of ≥20 cigarettes per day at 13-38 years of age. Survival time to heavy smoking was calculated as the number of years between initiation and the first assessment at which a study member smoked ≥20 cigarettes per day. | 155 | (25) |
| **Rapid Progression to Heavy Smoking.** Smoker of ≥20 cigarettes per day by 18 years of age. Most study members who ever became heavy smokers progressed to smoking ≥20 cigarettes per day by the 21 years of age assessment (63%). Therefore, we defined "rapid progression" as progression to smoking ≥20 cigarettes per day by the previous assessment at 18 years of age. | 61 | (10) |
| **Mature Phenotypes** | | |
| **Lifetime Cigarette Consumption (Pack-Years).** Pack-years = the number of cigarettes smoked per day divided by 20 and multiplied by the number of years smoked at that rate through 38 years of age.[38] The mean and SD of pack-years was calculated for ever-smokers. | 8.36 | (8.97) |
| **Nicotine Dependence.** Study members completed the FTND[41] at the age-21, -26, and -38 assessments. The FTND was developed to measure the construct of physical dependence and includes the facets of needing to smoke in the morning to alleviate overnight withdrawal, needing to smoke many cigarettes per day, and invariance in smoking behavior, eg, in the face of illness.[42] The FTND produces a score ranging from 0-10. Nicotine dependence is defined as a score ≥4.[43,44] We calculated survival time as years between smoking initiation and the first assessment at which a study member was nicotine dependent. Percent reflects ever-smokers who became nicotine dependent. | 170 | (27) |
| **Smoking to Cope With Stress.** Study members were interviewed about how they coped with stress associated with their relationships, work, and finances at 32 and 38 years of age. Study members rated the extent to which they used different coping strategies. One of these strategies was smoking more. Ratings of smoking as a coping strategy were averaged and the average was standardized to produce a score with a mean of 0 and an SD of 1. | 0 | (1) |
| **Cessation at 18-32 Years of Age.** To assess cessation failure before the period covered by the life history calendar, we used study member's reports collected at 18, 21, 26, and 32 years of age of whether they had made a quit attempt in the past year and whether they had failed in their quit attempt within one month. The number reflects study members who attempted cessation at least once. The percentage is calculated from study members who were ever daily smokers between ages 18 and 32 years (n=405). | 364 | (90) |
| **Cessation at 32-38 Years of Age.** At the assessment at 38 years of age, study members completed life history calendars[45] detailing their smoking behavior during each month from 32 to 38 years of age. Embedding recall of smoking behavior in a life-history calendar improves accuracy.[46] We used these data to identify 2 phenotypes: Relapses were ≥1 months of abstinence followed by ≥1 months of daily smoking; successful smoking cessation was abstinence for ≥1 year through the time of the interview. The number reflects study members who quit for ≥1 month between 32 and 38 years of age. The percentage is calculated from study members who became daily smokers by 32 years of age and were daily smokers for ≥1 month during 32-38 years of age (n=277). | 146 | (53) |

**Figure 2.** Smoking behavior in the Dunedin cohort. A, Developmental progression of smoking behavior in the Dunedin cohort. Study members reported their smoking status during in-person assessments at the ages of 11 (percentage of ever-smokers = 7%), 13 (13%), 15 (62%), 18 (66%), 21 (70%), 26 (70%), 32 (71%), and 38 (71%) years and their daily cigarette consumption at the ages of 13 (percentage of daily smokers = 1%), 15 (14%), 18 (31%), 21 (34%), 26 (35%), 32 (30%), and 38 (20%) years. We assessed nicotine dependence using the Fagerström Test of Nicotine Dependence (FTND),[41] completed by study members at the ages of 21, 26, and 38 years. We assessed cessation failure using study members' reports of quit attempts and outcomes at the ages of 18, 21, 26, 32, and 38 years. B, Measurements of developmental and mature smoking phenotypes. Data are number (percentage) of study members unless otherwise indicated.

Dunedin cohort members who initiated smoking at some point during follow-up (Figure 2).

## GENETIC RISK AND THE PROGRESSION OF SMOKING BEHAVIOR

Individuals at higher genetic risk were more likely to progress to smoking 20 or more cigarettes per day and did so more rapidly (HR = 1.35; 95% CI, 1.14-1.58). **Figure 3**A shows the cumulative hazards for smoking 20 cigarettes or more per day for individuals at low, average, and high genetic risk. An unexpected finding was that individuals who initiated smoking but who did not progress to daily smoking or to heavy smoking, so-called chippers, were at the lowest genetic risk of any group in the cohort (Figure 3B).

Among ever-smokers, 19% converted to daily smoking by age 15 years (early conversion) and 10% progressed to smoking 20 cigarettes or more per day by age 18 years (rapid progression to heavy smoking). Adolescents at higher genetic risk were more likely to convert to daily smoking early (RR = 1.24; 95% CI, 1.06-1.45) and to progress rapidly from smoking initiation to heavy smoking (RR = 1.43; 95% CI, 1.10-1.86).

Individuals at higher genetic risk accumulated more pack-years across 38 years of follow-up. Results from an

**Table. Effect Sizes for Genetic and Family History Associations With Developmental and Clinical Phenotypes of Smoking Behavior[a]**

| Variable | Effect Size Measure | Genetic Risk Score | Family History Score |
|---|---|---|---|
| Developmental phenotypes | | | |
| Smoking Initiation (among 880 individuals, 627 who ever initiated smoking) | | | |
| Ever-smoker status | RR | 0.98 (0.95 to 1.02) | 1.12 (1.07 to 1.17) |
| Lifetime hazard for smoking initiation | HR[b] | 1.01 (0.94 to 1.09) | 1.06 (0.98 to 1.15) |
| Progression from initiation to heavy smoking (among 627 ever-smokers) | | | |
| Early conversion to daily smoking (by the age of 15 years) | RR | 1.24 (1.06 to 1.45) | 1.52 (1.27 to 1.83) |
| Rapid progression to smoking ≥20 cigarettes per day (by the age of 18 years) | RR | 1.43 (1.10 to 1.86) | 1.68 (1.26 to 2.24) |
| Lifetime hazard for smoking ≥20 cigarettes per day | HR[b] | 1.35 (1.14 to 1.58) | 1.47 (1.23 to 1.76) |
| Mature phenotype | | | |
| Heavy smoking persistence (among 627 ever-smokers) | | | |
| Lifetime cigarette consumption (pack-years) | B | 1.05 (0.36 to 1.73) | 2.49 (1.80 to 3.19) |
| Count of assessments smoking ≥20 cigarettes per day | IRR | 1.26 (1.07 to 1.49) | 1.49 (1.24 to 1.80) |
| Nicotine dependence (among 627 ever-smokers) | | | |
| Lifetime hazard to becoming nicotine dependent (≥4 Fagerström symptoms) | HR[b] | 1.27 (1.09 to 1.47) | 1.53 (1.29 to 1.80) |
| Count of assessments with nicotine dependence | IRR | 1.22 (1.06 to 1.41) | 1.50 (1.28 to 1.75) |
| Smoking to cope with stress (ages 32-38 years, among 277 daily smokers) | | | |
| Smoking to cope score | B | 0.22 (0.11 to 0.32) | 0.09 (−0.06 to 0.24) |
| Cessation failure | | | |
| Ages of 18-32 years (405 daily smokers, 364 who attempted cessation) | | | |
| Risk of cessation failure | RR[b] | 1.11 (1.01 to 1.22) | 1.11 (1.00 to 1.23) |
| Ages of 32-38 years (277 daily smokers, 146 who quit for ≥1 mo) | | | |
| Hazard of relapse after quit attempts lasting ≥1 mo | HR[b] | 1.22 (1.02 to 1.45) | 0.96 (0.79 to 1.17) |
| Likelihood of successful cessation (among daily smokers) | RR | 0.73 (0.57 to 0.93) | 0.94 (0.73 to 1.20) |

Abbreviations: HR, hazard ratio; IRR, incident rate ratios; RR, relative risk.
[a] The correlation between the genetic risk score and the family history score was $r = 0.011$ ($P = .76$).
[b] Effect sizes were estimated from longitudinal data sets that included repeated observation of individuals over time.

ordinary least squares model indicated that each 1-unit increase in the GRS predicted an additional pack-year in lifetime cigarette consumption among ever-smokers (B = 1.05; 95% CI, 0.36-1.73) (**Figure 4**A). We also analyzed the persistence of heavy smoking as the number of assessments at which individuals smoked 20 cigarettes or more per day. Individuals at higher genetic risk smoked heavily at more assessments (incidence rate ratio [IRR] for number of assessments as a heavy smoker = 1.26; 95% CI, 1.07-1.49).

## GENETIC RISK AND NICOTINE DEPENDENCE

Through age 38 years, 27% of ever-smokers developed nicotine dependence. Individuals at higher genetic risk were more likely to become nicotine dependent compared with individuals at lower genetic risk and were nicotine dependent at more assessments (HR for nicotine dependence = 1.27; 95% CI, 1.09-1.47; IRR for assessments with nicotine dependence = 1.22; 95% CI, 1.06-1.41) (Figure 4B).

In addition to testing genetic associations with nicotine dependence, we also asked whether cohort members at higher genetic risk were more reliant on smoking to cope with stress. Among the 277 study members who smoked daily during ages 32 to 38 years, those at higher genetic risk relied more heavily on smoking as a coping strategy (B = 0.22; 95% CI, 0.11-0.32).

## GENETIC RISK AND SMOKING CESSATION

Assessment of cessation failure is challenging.[56] Therefore, we looked for convergent evidence across 2 approaches to testing genetic associations with cessation failure. We first analyzed study members' reports of cessation failure between the ages of 18 and 32 years. Across 14 years of follow-up, 405 cohort members smoked daily. A total of 90% of this group made at least one quit attempt, and 51% reported a cessation failure at 1 or more assessments. Cohort members at higher genetic risk were more likely to experience cessation failure in their quit attempts (RR = 1.11; 95% CI, 1.01-1.22).

We next used the month-to-month life-history calendars to look closely at cohort members' smoking behavior during their 30s, when cessation was most common. Across 72 months of follow-up, 277 cohort members smoked daily, and 53% of these smokers made a quit attempt lasting 1 month or more. Relapse was common (occurring in 62% of quitters). Quitters at higher genetic risk were more likely to relapse and did so sooner after quitting (HR = 1.22; 95% CI, 1.02-1.45). Only 20% of daily smokers achieved successful cessation (abstinent for ≥1 year through age 38 years). Smokers at higher genetic

risk were less likely to have achieved successful cessation at the end of follow-up (RR = 0.73; 95% CI, 0.57-0.93) (Figure 4C).

## GENETIC RISK, DEVELOPMENTAL PHENOTYPES OF SMOKING BEHAVIOR, AND ADULT SMOKING PROBLEMS

We derived an index of adult smoking problems from a principal components analysis of 3 indicators: (1) pack-years smoked by age 38 years, (2) total number of Fagerström Test of Nicotine Dependence symptoms across assessments, and (3) the number of assessments at which study members reported cessation failure. The adult smoking problems factor explained 78% of the variance in the 3 indicators. Individuals at higher genetic risk developed more smoking problems in adulthood ($r$ = 0.10, $P$ = .01). We next tested whether this association was accounted for by the more rapid developmental progression of smoking behavior among individuals at higher genetic risk. A total of 81% of this association was accounted for by the 2 adolescent developmental phenotypes of early conversion to daily smoking and rapid progression to smoking 20 or more cigarettes per day (eTable 4). As a further attempt to address the question of whether preventing rapid progression from smoking initiation to heavy smoking could mitigate genetic risks, we conducted a utopian control analysis.[57] We asked whether genetic risks continued to predict adult smoking problems in the subset of individuals who initiated smoking but who did not exhibit either of the rapid progression phenotypes (n = 454). In this subgroup, genetic risk was uncoupled from the development of smoking problems in adulthood ($r$ = 0.05, $P$ = .18).

## OVERLAP OF MOLECULAR GENETIC RISK AND FAMILY HISTORY OF SMOKING BEHAVIOR

The family history score and the GRS were uncorrelated ($r$ = 0.011). Both family history and the GRS predicted study members' smoking phenotypes (Table). When family history and the GRS were standardized and included in regression models simultaneously, the GRS and family history coefficients were unchanged and remained statistically significant (ie, genetic risk and family history were independent and additive predictors of smoking phenotypes). In the mediation analyses, adjustment for family history did not change results. Thus, the GRS contained different information about risk for developmental and mature phenotypes of smoking behavior compared with family history.

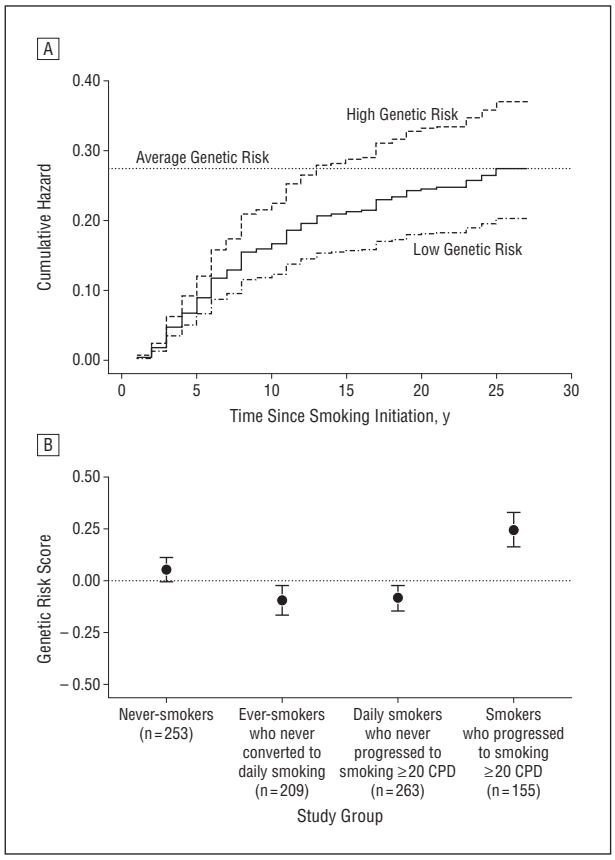



**Figure 3.** Genetic risk score (GRS) derived from genome-wide association study of smoking quantity is associated with the developmental progression of smoking behavior in a birth cohort of European-descent individuals. A, Individuals at higher genetic risk progressed more rapidly from smoking initiation to heavy smoking. This panel graphs hazard functions for onset of heavy smoking among individuals at low genetic risk (GRS = −1), average genetic risk (GRS = 0), and high genetic risk (GRS = 1). The dashed gray line marks the cumulative hazard for individuals at average genetic risk. The hazard function was estimated from a Cox proportional hazard model with time since onset of ever-smoking as the exposure time and the first assessment a study member reported smoking 20 or more cigarettes per day (CPD) as the failure event. The hazard model included all individuals who ever initiated smoking (n = 627). Individuals at higher genetic risk progressed more rapidly from smoking initiation to smoking 20 or more CPD (hazard ratio = 1.35; 95% CI, 1.14-1.58). B, Genetic risk was highest among individuals who progressed to heavy smoking and lowest among individuals who initiated smoking but who did not progress to heavy smoking. This panel shows the GRSs (±1 SE) for each group. A GRS of 0 corresponds to the average genetic risk in the cohort. Error bars reflect SEs of the subgroup means.

## COMMENT

Etiologic research on substance abuse highlights the importance of progression from initiation to heavy use during adolescence in the development of dependence in adulthood.[58,59] In this study, we linked the developmental progression of smoking behavior to genetic risk. We derived a GRS from GWASs of smoking quantity. This GRS was not related to smoking initiation. In fact, daily smokers who did not progress to heavy use were at lower genetic risk than individuals who never smoked. Among individuals who initiated smoking, those at higher genetic risk progressed more rapidly to heavy smoking and nicotine dependence, were more likely to become persistent heavy smokers and persistently nicotine dependent, and had more difficulty quitting. Critically, high genetic risk led individuals to become persistent heavy smokers, nicotine dependent, and unable to quit only to the extent that they progressed rapidly from smoking initiation to heavy smoking during adolescence.

The GWASs from which we derived our measure of genetic risk were designed to discover genetic correlates of smoking quantity. Therefore, the fact that genetic risks discovered by these GWASs do not predict smoking initiation is not entirely unexpected. Neverthe-

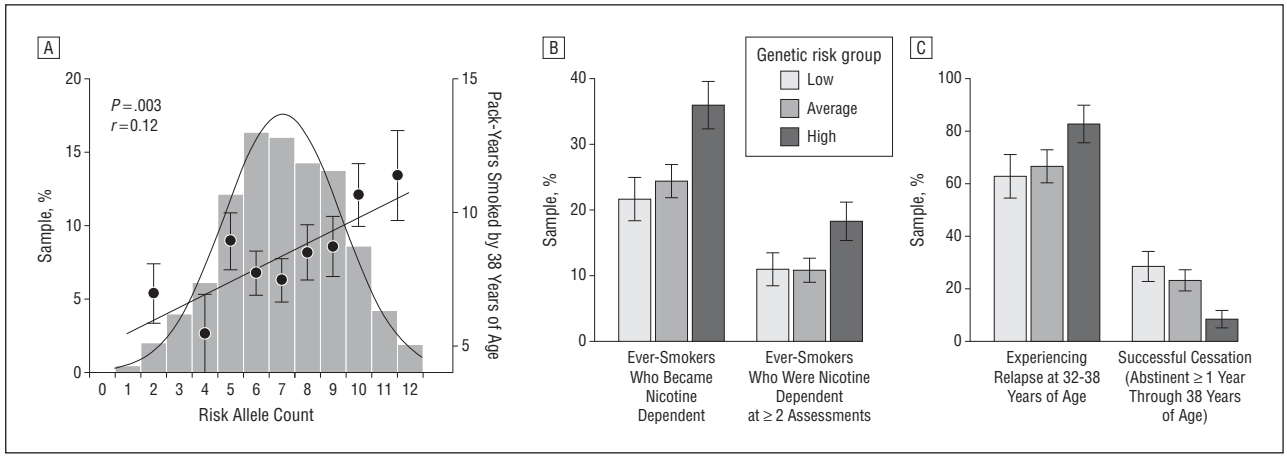



**Figure 4.** Genetic risk predicts mature phenotypes of smoking behavior. A, Among individuals who initiated smoking, those at higher genetic risk smoked more cigarettes by 38 years of age. Ever-smokers were all individuals who initiated smoking by 38 years of age (n = 627). The bars of the histogram graph the percentages of the sample carrying 1 to 12 risk alleles. The dots and SE bars reflect mean lifetime cigarette consumption (in pack-years) for ever-smokers carrying 1 to 3, 4, 5, 6, 7, 8, 9, 10, and 11 to 12 risk alleles. The regression line shows the association between the genetic risk score (GRS) and pack-years smoked by 38 years of age (Pearson correlation $r = 0.12$, $P = .003$). B, Ever-smokers at higher genetic risk were more likely to be nicotine dependent. The bars of the chart graph the proportion of ever-smokers at low (n = 157), average (n = 292), and high (n = 178) genetic risk who became nicotine dependent (≥4 Fagerström symptoms) by 38 years of age and who were nicotine dependent at 2 or more assessments. C, Smokers at higher genetic risk were more likely to experience cessation failure during their 30s. The bars of the chart graph the proportions of daily smokers at low, average, and high genetic risk who experienced relapse after a quit attempt lasting 1 month or longer and who achieved successful cessation (abstinence ≥1 year) through 38 years of age. Percentage with relapse was calculated from cohort members who quit smoking for 1 month or longer during 32 to 38 years of age (n = 36 for the low genetic risk group, n = 61 for the average genetic risk group, and n = 34 for the high genetic risk group). Percentage with successful cessation was calculated for cohort members who smoked daily during their 30s (n = 65 for the low genetic risk group, n = 120 for the average genetic risk group, and n = 77 for the high genetic risk group). B and C, Low genetic risk individuals had GRSs more than 0.5 SD below the cohort mean, average genetic risk individuals had GRSs within 0.5 SD of the cohort mean, and high genetic risk individuals had GRSs more than 0.5 SD above the cohort mean. Error bars reflect SEs.

less, that so-called chippers (light but persistent smokers)[60] in our cohort had below average genetic risk is consistent with the theory that the genetic risks captured in our score influence response to nicotine, not the propensity to initiate smoking.[17,61] Thus, our result affirms the value of using former and light smokers as a comparison group to heavy and nicotine dependent smokers in discovery analyses targeting these risks.

Previous research has related polymorphisms in the genes included in our genetic risk score to developmental phenotypes of smoking behavior[24,26,32-35] and to mature phenotypes of adult smoking problems.[29-31,62-64] To our knowledge, ours is the first study to track the relations of particular genetic risk variants with the development of smoking behavior from initiation through conversion to daily smoking and progression to heavy smoking and on to the mature phenotypes of persistent of heavy smoking, nicotine dependence, and struggles with cessation through midlife. Moreover, this extended follow-up allowed us to find, for the first time, that GWAS-identified variation in 15q25.1 and 19q13.2 influences adult smoking problems through a pathway mediated by adolescent progression from smoking initiation to heavy smoking. Our study is also the first, to our knowledge, to find that GWAS-identified SNPs provide information about smoking risks that cannot be ascertained from a family history, including information about risk for cessation failure.

These findings should be considered in light of 3 limitations. First, although the Dunedin Study sample consisted of European-descent individuals, as did the samples analyzed in the GWASs used to develop the GRS, we cannot rule out the possibility of population stratification. Further, replication in other populations is needed.[65] Second, our analyses of cessation were subject to censored data. The life-history calendars ended at the age of 38 years, and thus these data do not reflect relations with phenotypic events occurring after this age. In addition, self-reports of temporally remote events could be inaccurate because of forgetting or other biases. Third, the 4 decades of follow-up in the Dunedin Study coincided with major secular events, such as bans against smoking in the workplace. Comparisons of cohorts born at different times might elucidate gene-policy interactions in smoking behavior and speak to the generalizability of the current findings.[66,67]

Despite these limitations, this study has implications for etiologic research and public health. With respect to etiology, our study makes 3 contributions. First, next-generation sequencing studies and other efforts to ascertain causal variants responsible for GWAS signals may maximize their discovery potential by focusing on samples of young people strategically selected to reflect important developmental transitions. Such work could use experimental designs to test hypotheses about mechanisms of genetic risk on postinitiation phenotypes. Second, we demonstrated that a GRS based on the assumption of additive risks can be used to follow up GWAS results in a birth cohort far smaller than the original discovery samples. Future etiologic research can use GRSs to apply GWAS results to longitudinal studies. Third, results are consistent with the hypothesis in pediatric medicine that some adolescents, after only experimental use, are prone to quickly become heavy users and dependent.[68] This finding suggests that gene-environment interaction analyses of smoking and nicotine dependence may profit from a focus on environments that coincide with or immediately precede the adolescent period and influ-

ence the propensity of children at high genetic risk to initiate smoking. Smoking by peers is one such environment.[69] Tobacco control policies targeting youth may be another.[70,71]

Turning to public health, our research adds a genetic dimension to long-standing arguments that early prevention could be a critical strategy in reducing cigarette consumption.[72] Specifically, our findings and others'[32] suggest that initiatives that disrupt the developmental progression of smoking behavior, such as surtaxes and age restrictions on tobacco purchases, may ameliorate some genetic risks.[73] Moving beyond population-level prevention, we found that information about smoking risk captured in a score composed of GWAS-identified variants was independent of information that could be derived from a family history of smoking behavior. This novel finding suggests that genetic information could be used to identify high-risk youngsters for targeted prevention.[68,74] However, the associations we detected between the GRS and smoking phenotypes were small in magnitude. Small effect sizes do not preclude public health relevance,[75] but they caution against the use of genetic information to evaluate risk in individuals[76]; children who our study would classify at high genetic risk are not guaranteed to become addicted if they try smoking, and, even more importantly, children we would classify at low genetic risk are not immune to addiction. The public health use of the current findings must be tempered with recognition that most risk-associated genetic variation does not determine poor health outcomes, and, correspondingly, its absence does not guarantee protection.[77,78]

**Submitted for Publication:** March 9, 2012; final revision received July 12, 2012; accepted September 17, 2012.
**Published Online:** March 27, 2013. doi:10.1001/jamapsychiatry.2013.736
**Author Affiliations:** Department of Health Policy and Management, Gillings School of Public Health (Drs Belsky and Biddle), and Department of Genetics, School of Medicine (Dr Evans), University of North Carolina–Chapel Hill; Center for the Study of Aging and Human Development (Dr Belsky) and Department of Psychiatry and Behavioral Sciences (Drs Belsky, Moffitt, Houts, Meier, Sugden, and Caspi; Ms Harrington; and Mr Williams), Duke University Medical Center, and Institute for Genome Sciences and Policy (Drs Belsky, Moffitt, Houts, Meier, Sugden, and Caspi; Ms Harrington; and Mr Williams) and Department of Psychology and Neuroscience (Drs Belsky, Moffitt, Houts, Meier, Sugden, and Caspi; Ms Harrington; and Mr Williams), Duke University, Durham, North Carolina; Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, United Kingdom (Drs Moffitt, Sugden, and Caspi and Mr Williams); Center for Tobacco Research and Intervention, University of Wisconsin, and Department of Medicine, University of Wisconsin-Madison School of Medicine and Public Health, Madison (Dr Baker); and Dunedin Multidisciplinary Health and Development Research Unit, University of Otago, Otago, New Zealand (Dr Poulton).
**Correspondence:** Avshalom Caspi, PhD, Duke University/Duke University Medical Center, Center for Aging and the Study of Human Development, 2020 W Main St, Ste 201, Durham, NC 27708 (ac115@duke.edu).
**Conflict of Interest Disclosures:** None reported.
**Funding/Support:** This research received support from grant AG032282 from the US National Institute on Aging, grant MH077874 from the US National Institute on Mental Health, and grant G0101483 from the UK Medical Research Council. The Dunedin Multidisciplinary Health and Development Research Unit is supported by The New Zealand Health Research Council. Dr Belsky was supported in part by grant 1R36HS020524-01 from the US Agency for Healthcare Research and Quality and grant T32-AG000029 from the National Institute on Aging. Dr Baker was supported in part by grant 1K05CA139871 from the US National Cancer Institute. Dr Meier was supported in part by grant P30DA023026 from the US National Institute on Drug Abuse. Additional support was provided by the Jacobs Foundation.
**Online-Only Material:** The eMethods and eTables are available at http://www.jamapsych.com.
**Additional Contributions:** We thank the Dunedin Study members, their families, unit research staff, and study founder Phil Silva.

## REFERENCES

1. Centers for Disease Control and Prevention (CDC). Smoking-attributable mortality, years of potential life lost, and productivity losses—United States, 2000-2004. *MMWR Morb Mortal Wkly Rep*. 2008;57(45):1226-1228.
2. Centers for Disease Control and Prevention (CDC). Current cigarette smoking prevalence among working adults—United States, 2004-2010. *MMWR Morb Mortal Wkly Rep*. 2011;60(38):1305-1309.
3. Joseph AM, Fu SS, Lindgren B, et al. Chronic disease management for tobacco dependence: a randomized, controlled trial. *Arch Intern Med*. 2011;171(21):1894-1900.
4. Carpenter MJ, Hughes JR, Gray KM, Wahlquist AE, Saladin ME, Alberg AJ. Nicotine therapy sampling to induce quit attempts among smokers unmotivated to quit: a randomized clinical trial. *Arch Intern Med*. 2011;171(21):1901-1907.
5. Carson KV, Brinn MP, Labiszewski NA, Esterman AJ, Chang AB, Smith BJ. Community interventions for preventing smoking in young people *Cochrane Database Syst Rev*. 2011;(7):CD001291. doi:10.1002/14651858.CD001291.pub2.
6. Xian H, Scherrer JF, Madden PA, et al. The heritability of failed smoking cessation and nicotine withdrawal in twins who smoked and attempted to quit. *Nicotine Tob Res*. 2003;5(2):245-254.
7. Thorgeirsson TE, Gudbjartsson DF, Surakka I, et al; ENGAGE Consortium. Sequence variants at *CHRNB3-CHRNA6* and *CYP2A6* affect smoking behavior. *Nat Genet*. 2010;42(5):448-453.
8. Liu JZ, Tozzi F, Waterworth DM, et al; Wellcome Trust Case Control Consortium. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet*. 2010;42(5):436-440.
9. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet*. 2010;42(5):441-447.
10. Sullivan PF, Kendler KS. The genetic epidemiology of smoking. *Nicotine Tob Res*. 1999;1(suppl 2):S51-S57.
11. Baker TB, Weiss RB, Bolt D, et al. Human neuronal acetylcholine receptor A5-A3-B4 haplotypes are associated with multiple nicotine dependence phenotypes. *Nicotine Tob Res*. 2009;11(7):785-796.
12. Munafò MR, Johnstone EC, Walther D, Uhl GR, Murphy MF, Aveyard P. *CHRNA3* rs1051730 genotype and short-term smoking cessation. *Nicotine Tob Res*. 2011;13(10):982-988.
13. Lerman C, Jepson C, Wileyto EP, et al. Genetic variation in nicotine metabolism predicts the efficacy of extended-duration transdermal nicotine therapy. *Clin Pharmacol Ther*. 2010;87(5):553-557.
14. King DP, Paciga S, Pickering E, et al. Smoking cessation pharmacogenetics: analysis of varenicline and bupropion in placebo-controlled clinical trials. *Neuropsychopharmacology*. 2012;37(3):641-650.
15. Baker TB, Conti DV, Moffitt TE, Caspi A. *The Nicotine Dependence Phenotype: Translating Theoretical Perspectives and Extant Data Into Recommendations for Genetic Mapping*. Bethesda, MD: National Cancer Insititute; 2009.
16. Gelernter J. Developmental perspective on the role of genes in smoking risk. *Biol Psychiatry*. 2011;69(7):616-617.
17. Bierut LJ. Nicotine dependence and genetic variation in the nicotinic receptors. *Drug Alcohol Depend*. 2009;104(suppl 1):S64-S69.

18. Breslau N, Peterson EL. Smoking cessation in young adults: age at initiation of cigarette smoking and other suspected influences. *Am J Public Health*. 1996; 86(2):214-220.
19. Buchmann AF, Blomeyer D, Jennen-Steinmetz C, et al. Early smoking onset may promise initial pleasurable sensations and later addiction [published online ahead of print October 4, 2011]. *Addict Biol*. doi:10.1111/j.1369-1600.2011.00377.x.
20. Kandel DB, Hu MC, Griesler PC, Schaffran C. On the development of nicotine dependence in adolescence. *Drug Alcohol Depend*. 2007;91(1):26-39.
21. Dierker L, Swendsen J, Rose J, He J, Merikangas K. Transitions to regular smoking and nicotine dependence in the Adolescent National Comorbidity Survey (NCS-A). *Ann Behav Med*. 2012;43(3):394-401.
22. Moffitt TE, Caspi A, Rutter M, Silva PA. *Sex Differences in Antisocial Behavior: Conduct Disorder, Delinquency, and Violence in the Dunedin Longitudinal Study*. Cambridge, England: Cambridge University Press; 2001.
23. Sullivan PF. Spurious genetic associations. *Biol Psychiatry*. 2007;61(10):1121-1126.
24. Ducci F, Kaakinen M, Pouta A, et al. *TTC12-ANKK1-DRD2* and *CHRNA5-CHRNA3-CHRNB4* influence different pathways leading to smoking behavior from adolescence to mid-adulthood. *Biol Psychiatry*. 2011;69(7):650-660.
25. Wassenaar CA, Dong Q, Wei Q, Amos CI, Spitz MR, Tyndale RF. Relationship between *CYP2A6* and *CHRNA5-CHRNA3-CHRNB4* variation and smoking behaviors and lung cancer risk. *J Natl Cancer Inst*. 2011;103(17):1342-1346.
26. Rodriguez S, Cook DG, Gaunt TR, Nightingale CM, Whincup PH, Day IN. Combined analysis of *CHRNA5*, *CHRNA3* and *CYP2A6* in relation to adolescent smoking behaviour. *J Psychopharmacol*. 2011;25(7):915-923.
27. Vrieze SI, McGue M, Iacono WG. The interplay of genes and adolescent development in substance use disorders: Leveraging findings from GWAS meta-analyses to test developmental hypotheses about nicotine consumption. *Hum Genet*. 2012;131(5):791-801.
28. Plomin R, Haworth CMA, Davis OSP. Common disorders are quantitative traits. *Nat Rev Genet*. 2009;10(12):872-878.
29. Berrettini WH, Doyle GA. The *CHRNA5-A3-B4* gene cluster in nicotine addiction. *Mol Psychiatry*. 2012;17(9):856-866.
30. Malaiyandi V, Sellers EM, Tyndale RF. Implications of *CYP2A6* genetic variation for smoking behaviors and nicotine dependence. *Clin Pharmacol Ther*. 2005; 77(3):145-158.
31. Kubota T, Nakajima-Taniguchi C, Fukuda T, et al. *CYP2A6* polymorphisms are associated with nicotine dependence and influence withdrawal symptoms in smoking cessation. *Pharmacogenomics J*. 2006;6(2):115-119.
32. Weiss RB, Baker TB, Cannon DS, et al. A candidate gene approach identifies the *CHRNA5-A3-B4* region as a risk factor for age-dependent nicotine addiction. *PLoS Genet*. 2008;4(7):e1000125.
33. Audrain-McGovern J, Al Koudsi N, Rodriguez D, Wileyto EP, Shields PG, Tyndale RF. The role of *CYP2A6* in the emergence of nicotine dependence in adolescents. *Pediatrics*. 2007;119(1):e264-e274.
34. O'Loughlin J, Paradis G, Kim W, et al. Genetically decreased *CYP2A6* and the risk of tobacco dependence: a prospective study of novice smokers. *Tob Control*. 2004;13(4):422-428.
35. Huang S, Cook DG, Hinks LJ, et al. *CYP2A6*, *MAOA*, *DBH*, *DRD4*, and *5HT2A* genotypes, smoking behaviour and cotinine levels in 1518 UK adolescents. *Pharmacogenet Genomics*. 2005;15(12):839-850.
36. Saccone NL, Culverhouse RC, Schwantes-An TH, et al. Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. *PLoS Genet*. 2010;6(8):e1001053. doi:10.1371/journal.pgen.1001053.
37. GENEVA ARIC Project. *Quality Control Report for the ARIC GWAS Database*. Bethesda, MD: The National Institutes of Health Database of Genotypes and Phenotypes (dbGaP); 2009. dbGaP accession No. phs000090.v1.p1 ed.
38. GENEVA Alcohol Dependence Project. *Quality Control Report for the SAGE GWAS*. Bethesda, MD: National Institutes of Health Database of Genotypes and Phenotypes (dbGaP); 2008. dbGaP accession No. phs000092.v1.p1 ed.
39. Poulton R, Caspi A, Milne BJ, et al. Association between children's experience of socioeconomic disadvantage and adult health: a life-course study. *Lancet*. 2002; 360(9346):1640-1645.
40. Milne BJ, Moffitt TE, Crump R, et al. How should we construct psychiatric family history scores? a comparison of alternative approaches from the Dunedin Family Health History Study. *Psychol Med*. 2008;38(12):1793-1802.
41. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. *Br J Addict*. 1991;86(9):1119-1127.
42. Fagerström KO. Measuring degree of physical dependence to tobacco smoking with reference to individualization of treatment. *Addict Behav*. 1978;3(3-4):235-241.
43. Bierut LJ, Madden PA, Breslau N, et al. Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet*. 2007; 16(1):24-35.
44. Saccone SF, Hinrichs AL, Saccone NL, et al. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet*. 2007;16(1):36-49.
45. Caspi A, Moffitt TE, Thornton A, et al. The life history calendar: a research and clinical assessment method for collecting retrospective event-history data. *Int J Method Psych*. 1996;6(2):101-114.
46. Belli RF, Smith LM, Andreski PM, Agrawal S. Methodological comparisons between CATI event history calendar and standardized conventional questionnaire instruments. *Public Opin Q*. 2007;71(4):603-622.
47. McNutt LA, Wu C, Xue X, Hafner JP. Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am J Epidemiol*. 2003;157(10): 940-943.
48. Gardner W, Mulvey EP, Shaw EC. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychol Bull*. 1995; 118(3):392-404.
49. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*. 1988;44(4):1049-1060.
50. Prentice RL, Williams BJ, Peterson AV. On the regression-analysis of multivariate failure time data. *Biometrika*. 1981;68(2):373-379.
51. Mackinnon DP, Dwyer JH. Estimating mediated effects in prevention studies. *Eval Rev*. 1993;17(2):144-158.
52. Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav Res Methods*. 2008; 40(3):879-891.
53. Preacher KJ, Kelley K. Effect size measures for mediation models: quantitative strategies for communicating indirect effects. *Psychol Methods*. 2011;16(2): 93-115.
54. Preacher KJ, MacCallum RC. Repairing Tom Swift's electric factor analysis machine. *Underst Stat*. 2003;2(1):13-43.
55. *Stata/MP for Windows* [computer program]. Version 11.1. College Station, TX: Stata Corp LP; 2010.
56. Hughes JR, Callas PW. Definition of a quit attempt: a replication test. *Nicotine Tob Res*. 2010;12(11):1176-1179.
57. Lubinski D. Cognitive epidemiology: with emphasis on untangling cognitive ability and socioeconomic status. *Intelligence*. 2009;37(6):625-633.
58. Jackson KM, Sher KJ, Rose RJ, Kapiro J. *Trajectories of Tobacco Use From Adolescence to Adulthood: Are the Most Informative Phenotypes Tobacco Specific?* Bethesda, MD: National Cancer Institute; 2009.
59. Bierut LJ. Genetic vulnerability and susceptibility to substance dependence. *Neuron*. 2011;69(4):618-627.
60. Shiffman S. Tobacco "chippers": individual differences in tobacco dependence. *Psychopharmacology (Berl)*. 1989;97(4):539-547.
61. Greenbaum L, Lerer B. Differential contribution of genetic variation in multiple brain nicotinic cholinergic receptors to nicotine dependence: recent progress and emerging open questions. *Mol Psychiatry*. 2009;14(10):912-945.
62. Broms U, Wedenoja J, Largeau MR, et al. Analysis of detailed phenotype profiles reveals *CHRNA5-CHRNA3-CHRNB4* gene cluster association with several nicotine dependence traits. *Nicotine Tob Res*. 2012;14(6):720-733.
63. Benowitz NL. Nicotine addiction. *N Engl J Med*. 2010;362(24):2295-2303.
64. Chen LS, Baker TB, Piper ME, Breslau N, Cannon DS, Doheny KF, Gogarten SM, Johnson EO, Saccone NL, Wang JC, Weiss RB, Goate AM, Bierut LJ. Interplay of genetic risk factors (*CHRNA5-CHRNA3-CHRNB4*) and cessation treatments in smoking cessation success *Am J Psychiatry*. 2012;169(7):735-742.
65. Saccone NL, Schwantes-An TH, Wang JC, et al. Multiple cholinergic nicotinic receptor genes affect nicotine dependence risk in African and European Americans. *Genes Brain Behav*. 2010;9(7):741-750.
66. Boardman JD, Blalock CL, Pampel FC, Hatemi PK, Heath AC, Eaves LJ. Population composition, public policy, and the genetics of smoking. *Demography*. 2011; 48(4):1517-1533.
67. Boardman JD, Blalock CL, Pampel FC. Trends in the genetic influences on smoking. *J Health Soc Behav*. 2010;51(1):108-123.
68. DiFranza JR, Richmond JB. Let the children be heard: lessons from studies of the early onset of tobacco addiction. *Pediatrics*. 2008;121(3):623-624.
69. Johnson EO, Chen LS, Breslau N, et al. Peer smoking and the nicotinic receptor genes: an examination of genetic and environmental risks for nicotine dependence. *Addiction*. 2010;105(11):2014-2022.
70. Boardman JD. State-level moderation of genetic tendencies to smoke. *Am J Public Health*. 2009;99(3):480-486.
71. Boardman JD, Saint Onge JM, Haberstick BC, Timberlake DS, Hewitt JK. Do schools moderate the genetic determinants of smoking? *Behav Genet*. 2008;38(3): 234-246.
72. Galea S, Nandi A, Vlahov D. The social epidemiology of substance use. *Epidemiol Rev*. 2004;26:36-52.
73. Fiore MC, Baker TB. Stealing a march in the 21st century: accelerating progress in the 100-year war against tobacco addiction in the United States. *Am J Public Health*. 2009;99(7):1170-1175.
74. Rose G. Sick individuals and sick populations. *Int J Epidemiol*. 1985;14(1):32-38.
75. Smith GD. Epidemiology, epigenetics and the 'Gloomy Prospect': embracing randomness in population health research and practice. *Int J Epidemiol*. 2011; 40(3):537-562.
76. Ioannidis JPA. Personalized genetic prediction: too limited, too expensive, or too soon? *Ann Intern Med*. 2009;150(2):139-141.
77. Evans JP, Meslin EM, Marteau TM, Caulfield T. Genomics. Deflating the genomic bubble. *Science*. 2011;331(6019):861-862.
78. Guttmacher AE, Porteous ME, McInerney JD. Educating health-care professionals about genetics and genomics. *Nat Rev Genet*. 2007;8(2):151-157.

# Supplementary Online Content

Belsky DW, Moffitt TE, Baker TB, et al. Polygenic risk and the developmental progression to heavy, persistent smoking and nicotine dependence [published online March 27, 2013]. *JAMA Psychiatry*. doi:10.1001/jamapsychiatry.2013.736.

**eAppendix 1.** Search strategy for MEDLINE (using PubMed).
**eAppendix 2.** Search strategy for EMBASE (using Embase.com).
**eFigure.** Flow of information through the different phases of the review
**eTable 1.** Diagnostic performance of serologic tests: test combinations
**eTable 2.** Results of quality assessment per study.
**eReferences**

This supplementary material has been provided by the authors to give readers additional information about their work.

Supplemental Materials to DW Belsky et al. “Polygenic risk accelerates the developmental progression to heavy, persistent smoking and nicotine dependence.”

## Supplemental Methods.

To select single-nucleotide polymorphisms (SNPs) for inclusion in our genetic risk score (GRS), we examined the results of 3 meta-analyses of genome-wide association studies (GWASs) of smoking quantity.[1-3] We first selected regions with genome-wide significant associations (*P*<.001) in at least 2 of the 3 meta-analyses. Next, we pruned the set of genome-wide significant SNPs using a linkage disequilibrium (LD) threshold of $R^2$=0.60. This procedure grouped SNPs into clusters on the basis of LD queried from the Hap Map v22 and 1000 Genomes databases (CEU samples) using the Broad Institute’s SNAP tool (http://www.broadinstitute.org/mpg/snap/ldsearch.php). From within each cluster, we selected the SNP with the lowest p-value in the meta-analysis by Thorgeirsson and colleagues[1] (the only meta-analysis to report exact p-values for all SNPs).

We then compared the set of 6 SNPs derived as described above to a set of available genotypes from an Illumina BeadPlex Array previously constructed to genotype SNPs identified in GWAS of obesity, asthma, and smoking. We selected the best available matches among SNPs passing quality control (Hardy Weinberg equilibrium >0.01; call rate of >95% in samples with valid calls for >95% of SNPs).

Supplemental Materials to DW Belsky et al. "Polygenic risk accelerates the developmental progression to heavy, persistent smoking and nicotine dependence."

**eTable 1. Replication of Association Between the Genetic Risk Score and Smoking Quantity in the Dunedin Cohort and the ARIC and SAGE Databases[a]**

| | Ordinary Least Squares | | Negative Binomial | |
|---|---|---|---|---|
| | B | 95% CI | IRR | 95% CI |
| **Dunedin (n=627)** | | | | |
| Full GRS | 0.08 | (0.03-0.14) | 1.16 | (1.05-1.29) |
| GRS excluding proxy for rs16969968 | 0.09 | (0.03-0.14) | 1.18 | (1.06-1.31) |
| GRS excluding proxy for rs6495308 | 0.09 | (0.04-0.15) | 1.19 | (1.07-1.32) |
| **ARIC (n=4,804)** | | | | |
| Full GRS | 0.08 | (0.05-0.12) | 1.07 | (1.05-1.10) |
| GRS excluding proxy for rs16969968 | 0.07 | (0.03-0.11) | 1.06 | (1.04-1.09) |
| GRS excluding proxy for rs6495308 | 0.07 | (0.05-0.10) | 1.07 | (1.04-1.09) |
| **SAGE (n=1,992)** | | | | |
| Full GRS | 0.06 | (0.02-0.11) | 1.09 | (1.04-1.15) |
| GRS excluding proxy for rs16969968 | 0.04 | (0.01-0.08) | 1.07 | (1.01-1.13) |
| GRS excluding proxy for rs6495308 | 0.05 | (0.01-0.10) | 1.08 | (1.03-1.14) |

[a]The dependent variable in the analyses is a 4-category measure of lifetime smoking quantity: 0-10 cigarettes/day; 11-20 cigarettes/day; 21-30 cigarettes/day; 31+ cigarettes/day. The categories match those used by the Fagerstrom Test of Nicotine Dependence.[4] Analyses tested the full genetic risk score (gray highlights) and versions of this score that excluded each of the two best-replicated GWAS hits for smoking phenotypes. Effects are presented for a 1-standard deviation increase in the genetic risk scores (GRSs). Two regression approaches were used to analyze the association between genetic risk and smoking quantity: ordinary least squares (OLS), which treated the 4-category smoking quantity measure as a continuous variable, and negative binomial regression, which treated the 4-category measure as a count variable. Regression coefficients (B) are presented from the OLS models. Incident rate ratios (IRRs) are presented from the negative binomial regression models. Analyses include European-descent individuals who ever smoked. The ARIC database included individuals sampled from 3 US regions (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000090.v1.p1). The SAGE database included individuals drawn from 3 separate studies (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000092.v1.p1). Analyses of these databases accounted for their

multi-part structure using fixed and random effects in the case of the OLS models and fixed effects in the case of the negative binomial models.

Supplemental Materials to DW Belsky et al. "Polygenic risk accelerates the developmental progression to heavy, persistent smoking and nicotine dependence."

**eTable 2. Single-Nucleotide Polymorphisms (SNPs) Included in the Genome Risk Score**[a]

| Genes | GWAS SNP | Alleles | Effect Allele | Frequency (HapMap), % | Dunedin SNP | LD With GWAS SNP | Alleles | Effect Allele | Frequency (Dunedin), % |
|---|---|---|---|---|---|---|---|---|---|
| *CHRNA5, CHRNA3, CHRNB4* | rs16969968 | A/G | A | 39 | rs10519203 | 0.93 | A/G | G | 34 |
| | rs6495308 | C/T | T | 80 | rs4887069 | 1.00 | A/G | A | 79 |
| *ADAMTS7, MORF4L1* | rs12595538 | A/T | A | 62 | rs7164529 | 0.90 | A/G | G | 61 |
| | rs8032771 | A/G | A | 52 | rs11072810 | 0.97 | C/T | T | 50 |
| *EGLN2* | rs7937 | C/T | T | 55 | rs7937 | 1.00 | C/T | T | 57 |
| *CYP2A6* | rs4105144 | A/G | G | 74 | rs8102683 | 0.87 | C/T | C | 73 |

[a]Effect allele frequencies for the genome-wide association study (GWAS) SNPs are based on the HapMap CEU sample (release 22 for SNPs rs12595538, rs8032771, and rs4105144; version 3 release 2 for SNPs rs16969968 and rs6495308). Linkage disequilibrium (LD in terms of $R^2$) was obtained from 1000 Genomes project data for all SNPs except rs4105144. LD between this SNP and rs8102683 was obtained using HapMap Release 22 data. All allele frequency and linkage queries were run through the Broad Institute's SNAP tool (http://www.broadinstitute.org/mpg/snap/ldsearch.php). Effect allele frequencies for the SNPs genotyped in the Dunedin sample are based on n=880 European-descent study members.

Supplemental Materials to DW Belsky et al. “Polygenic risk accelerates the developmental progression to heavy, persistent smoking and nicotine dependence.”

**eTable 3. Correlations Among Measures of Adult Smoking Problems[a]**

| | | (1) | (2) | (3) |
|---|---|---|---|---|
| (1) | Lifetime cigarette consumption (pack-years) | 1.00 | | |
| (2) | Total Fagerstr□m symptoms | 0.81 | 1.00 | |
| (3) | Total No. of cessation failures (relapse) | 0.52 | 0.51 | 1.00 |

[a]Pearson correlations (*r*) were calculated among n=627 ever smokers. Total Fagerstr□m Symptoms was the sum of Fagerstr□m Test of Nicotine Dependence (FTND) scores across all measurements. All correlations were statistically significant at *P*<.001.

Supplemental Materials to DW Belsky et al. “Polygenic risk accelerates the developmental progression to heavy, persistent smoking and nicotine dependence.”

**eTable 4. Associations Between Genetic Risk and Clinical Phenotypes of Smoking Behavior Are Mediated by Developmental Phenotypes of Rapid Progression From Smoking Initiation to Heavy Smoking[a]**

| Latent Adult Smoking Problems Factor | Total Effect of Genetic Risk | Direct (Unmediated) Effect of Genetic Risk | Indirect Effect of Genetic Risk Mediated Through Developmental Phenotypes | Proportion of Total Effect Accounted for by the Indirect Effect, % |
|---|---|---|---|---|
| B | 0.15 | 0.03 | 0.12 | 81 |
| 95% CI | 0.05 to 0.26 | -0.04 to 0.10 | 0.05 to 0.20 | |
| *P* value | <.001 | <.001 | <.001 | |

[a]The dependent variable in the mediation analysis was the latent smoking problems factor. Indirect, direct, and total effects were estimated from the structural equation described by MacKinnon and Dwyer implemented using the methods described by Preacher and colleagues. [5-7] Percentile-based 95% CIs were estimated from 1000 bootstrap repetitions. Developmental phenotypes were early conversion to daily smoking (by age 15 years) and rapid progression to heavy smoking (by age 18 years). Both developmental phenotypes were associated with the latent adult smoking problems factor and with the individual clinical phenotypes (*P*<.001 for all). Collectively, early conversion to daily smoking and rapid progression to heavy smoking explained 23% of the variance in the latent smoking problems factor.

Supplemental Materials to DW Belsky et al. “Polygenic risk accelerates the developmental progression to heavy, persistent smoking and nicotine dependence.”

**References**


1. Thorgeirsson TE, Gudbjartsson DF, Surakka I, et al. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet.* 2010;42(5):448-453.
2. Liu JZ, Tozzi F, Waterworth DM, Pillai SG, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet.* 2010;42(5):436-440.
3. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet.* 2010;42(5):441-447.
4. Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO. The Fagerstr□m Test for Nicotine Dependence: a revision of the Fagerstr□m Tolerance Questionnaire. *Br J Addict.* 1991;86(9):1119-1127.
5. Mackinnon DP, Dwyer JH. Estimating mediated effects in prevention studies. *Eval Rev.* 1993;17(2):144-158.
6. Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav Res Meth.* 2008;40(3):879-891.
7. Preacher KJ, Kelley K. Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychol Methods.* 2011;16(2):93-115.