



Original Contribution

Association of the Calcyon Neuron-Specific Vesicular Protein Gene (*CALY*) With Adolescent Smoking Initiation in China and California

Dalin Li, Stephanie J. London, Jinghua Liu, Wonho Lee, Xuejuan Jiang, David Van Den Berg, Andrew W. Bergen, Denise Nishita, Nahid Waleh, Gary E. Swan, Peggy Gallaher, Chih-Ping Chou, Jean C. Shih, Jennifer B. Unger, W. James Gauderman, Frank Gilliland, C. Anderson Johnson, and David V. Conti*

* Correspondence to Dr. David V. Conti, Division of Biostatistics, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, 1501 San Pablo Street, ZNI 101, MC 2821, Los Angeles, CA 90089 (e-mail: dconti@usc.edu).

Initially submitted March 19, 2010; accepted for publication December 3, 2010.

Although previous investigations have indicated a role for genetic factors in smoking initiation, the underlying genetic mechanisms are still unknown. In 2,339 adolescents from a Chinese Han population in the Wuhan Smoking Prevention Trial (Wuhan, China, 1998–1999), the authors explored the association of 57 genes in the dopamine pathway with smoking initiation. Using a conservative approach for declaring significance, positive findings were further examined in an independent sample of 603 Caucasian adolescents followed for up to 10 years as part of the Children’s Health Study (Southern California, 1993–2009). The authors identified 1 single nucleotide polymorphism (rs2298122) in the calcyon neuron-specific vesicular protein gene (*CALY*) that was positively associated with smoking initiation in females (odds ratio = 2.21, 95% confidence interval: 1.49, 3.27; $P = 8.4 \times 10^{-5}$) in the Wuhan Smoking Prevention Trial cohort, and they replicated the association in females from the Children’s Health Study cohort (hazard rate ratio = 2.05, 95% confidence interval: 1.27, 3.31; $P = 0.003$). These results suggest that the *CALY* gene may influence smoking initiation in adolescents, although the potential roles of underlying psychological characteristics that may be components of the smoking-initiation phenotype, such as impulsivity or novelty-seeking, remain to be explored.

adolescent; dopamine; genetic association studies; smoking

Abbreviations: *CALY*, calcyon neuron-specific vesicular protein gene; CHS, Children’s Health Study; CI, confidence interval; CRHR1, corticotropin-releasing hormone receptor 1 gene; SNP, single nucleotide polymorphism; WSPT, Wuhan Smoking Prevention Trial.

As the leading cause of preventable morbidity and mortality in the world, smoking results in 5 million premature deaths worldwide each year (1, 2), and this number will increase to more than 10 million by the year 2020 (3, 4). Previous research has indicated that many adult smokers started smoking in early adolescence (5), and early initiation of smoking greatly increases the likelihood of adult smoking dependence (6), failure to quit (7), and smoking-related health problems (8). Numerous studies have documented demographic, psychological, social, and cultural risk and protective factors for smoking initiation. Demographic factors include age, ethnicity, immigrant status, and acculturation (9, 10). Psychological factors include depression

(8, 11), impulsivity (12), and other personality traits. Social factors include the smoking behavior of friends, siblings, and parents (13, 14), bonding with one’s school or other prosocial institutions (15), family cohesion and conflict (16), and social norms. Cultural factors include cultural values, discrimination, and westernization (17). While it is clear that these factors influence smoking initiation, there is evidence supporting an important role of genetic factors as well (18, 19).

Although many previous investigations have focused on the genetic basis of nicotine dependence and smoking cessation (20–23), there has been less research on smoking initiation (24–27). Most candidate-gene-based association

studies on smoking initiation have focused on the effects of a limited number of genes, with a limited number of variants within each gene (28–32). Furthermore, 2 recent genome-wide association studies failed to identify single nucleotide polymorphisms (SNPs) associated with smoking initiation on the genome-wide significance level in adults (33, 34). In contrast to these 2 distinct study designs, a pathway-based study (35, 36) inspects the association of multiple variants in multiple genes in 1 or more plausible pathways based on a priori biologic information. By including multiple variants and multiple genes and focusing on hypothesis-driven rather than purely exploratory methods, a well-designed pathway-based investigation could offer some advantages for identifying variants and providing clues to the mechanisms underlying this complex behavior (37, 38).

There is ample evidence indicating the role of the dopamine pathway in smoking-related behaviors. The dopamine pathway plays a central role in the brain's reward system (39), potentially motivating drug-seeking and other behaviors related to addiction (40). Furthermore, the dopamine pathway is involved in many psychological traits that often co-occur with smoking, such as depression and anxiety (41–43). Moreover, previous investigations demonstrated that rats with high extracellular dopamine levels have persistently high exploratory activity (44) and that overexpression of the dopamine transporters can lead to impulsive behaviors (45). Both of these phenotypes can affect a person's reaction to smoking cues, with a subsequent impact on smoking initiation (46, 47).

With a cross-sectional sample of more than 2,300 adolescents from a genetically homogeneous Han population in Wuhan, China, we explored the association between smoking initiation and SNPs in 57 dopamine pathway genes. Significant findings were then examined in the Children's Health Study (CHS), an independent cohort of adolescents in Southern California, whose smoking behavior has been followed for up to 10 years.

MATERIALS AND METHODS

Wuhan Smoking Prevention Trial

Study sample and design. The Wuhan Smoking Prevention Trial (WSPT) is a longitudinal smoking prevention trial initiated by the China Center for Disease Control and Prevention and the Department of Preventive Medicine at the University of Southern California (Los Angeles, California) in 1998 (48). The cohort includes 2,661 adolescents from 14 middle schools in 7 urban districts in Wuhan, China, who were enrolled between December 1998 and January 1999. The participating adolescents completed a 200-item survey for assessment of their baseline smoking behavior and social and economic status, with a response rate of 97%. Buccal cells were collected for future DNA extraction (49). The questionnaires, study protocol, data collection, and DNA sample collection were approved by the institutional review boards at the University of Southern California, the China Center for Disease Control and Prevention, and the US National Institute of Environmental Health Sciences.

Phenotype assessment. Following the methods of large-scale investigations such as the Global Youth Tobacco Survey (50), smoking initiation was determined from participants' responses to a question on lifetime tobacco exposure ("Have you *ever* tried smoking, even a few puffs?"). Never smokers were defined as subjects who reported that they had never smoked a cigarette, not even a puff. Ever smokers were defined as students who reported that they had ever tried smoking. This measure was used instead of alternative measures such as past-month smoking or daily smoking because it is more sensitive to low-frequency smoking behavior in samples with a low prevalence of smoking, such as young adolescents.

Gene and marker selection. Details on the selection of candidate genes in the dopamine pathway are given elsewhere (51). Briefly, 57 candidate genes were selected on the basis of a systematic review of the pathway-based databases (52), as well as comprehensive bioinformatics literature mining. The selected candidate genes mainly consisted of genes encoding proteins involved in the dopamine-mediated reward pathway, including dopamine receptors, the dopamine transporter, and enzymes involved in biosynthesis and metabolism of dopamine. Genes hypothesized to play a role in modulating the dopamine pathway (e.g., genes in the serotonin and opioid pathways) were also included. Nine neuronal nicotinic acetylcholine receptor subunit genes which have been reported to regulate dopamine release in the brain (53) were also included. The complete list of selected candidate genes is shown in Web Table 1, which is posted on the *Journal's* Web site (<http://aje.oxfordjournals.org/>).

We selected 1,295 SNPs within our candidate genes. Several candidate genes are positioned adjacent to each other on a chromosome, reducing the regions of interest for tag SNP selection to 53 (51 gene regions on the autosomes and 2 on the X chromosome). SNP selection within each gene region (± 10 kilobases) was carried out with a beta version of the Snagger software (54), using HapMap data from the Han Chinese (CHB) and Caucasian (CEU) populations (Genome Build 35) (55). Among the 1,295 selected SNPs, 118 have an a priori putative function, and conditional on their inclusion, an additional 1,177 tag SNPs were chosen to capture the underlying genetic structure. Overall, within the 53 gene regions, 93% of all SNPs in HapMap have substantial linkage disequilibrium ($r^2 \geq 0.80$) with at least 1 SNP genotyped in this study in the HapMap Chinese population, indicating that the selected SNPs capture the genetic structure in these gene regions adequately.

Although the Chinese population is believed to be genetically homogenous, in order to identify unknown population substructure we also genotyped 233 ancestry informative markers (56, 57) in the WSPT.

Genotyping. SNP genotyping was performed in the Genomics Core Facility at the University of Southern California's Norris Comprehensive Cancer Center using the Illumina GoldenGate Genotyping Assay (Illumina, San Diego, California). Quality control procedures included automated protocols utilizing robotics and bar coding for the entire genotyping process and the inclusion of replicates and HapMap CEPH trios to identify genotyping errors.

Among the 2,661 subjects, genotype data were available for 2,443 persons. Of the 1,295 SNPs genotyped, 47 with a call rate of zero were excluded. Thereafter, 95 persons with call rates less than 95% were excluded. After removal of these persons, 25 SNPs with call rates less than 95% were excluded. An additional 9 persons with no phenotype information were removed, leaving 2,339 subjects available for analysis. Of the remaining 1,223 SNPs, 92% had a SNP call rate of 99% or more.

Imputation of untyped SNPs. Untyped SNPs were imputed on the basis of the haplotype structure of the HapMap East Asian population (58) for the 51 autosomal chromosome gene regions in the WSPT. Genotypes at 2,260 SNPs were imputed using MACH software (59). Of the imputed genotypes, 189 SNPs with imputation quality scores lower than 0.80 were excluded. After combining the genotyped and imputed SNPs, 165 SNPs with P values smaller than 0.005 from the exact Hardy-Weinberg equilibrium test and 429 SNPs with minor allele frequencies less than 0.01 were removed. Thus, 2,700 SNPs were included in the final analysis.

Population stratification analysis. The STRUCTURE program (60) was used to estimate the coefficient of ancestry for each person in the WSPT sample. No large-scale mixture was observed for any of the subjects (96.8% of the participants had an estimated Asian coefficient of ancestry greater than 0.90). Furthermore, the results were similar from analyses with and without adjustment for the coefficient of ancestry.

Statistical analysis. A generalized linear mixed model (with logistic link) was employed to evaluate the effect of each SNP on smoking initiation with school-level variation modeled through a school-specific intercept. Considering the previously documented gender difference in smoking behavior among Chinese adolescents (61), the effect of a given SNP was first analyzed with all subjects and then with data stratified by gender. Age and gender (when applicable) were included in the model as covariates.

To determine the heredity model, additive, dominant, and recessive models were tested, and the model with the smallest P value was selected. To obtain valid P values, we adjusted the final reported P value using the P_{ACT} (P value adjusted for correlated tests) approach (62). We first performed this adjustment to account for the determination of the heredity model and then performed the adjustment separately in each gene region to account for the multiple correlated tests due to linkage disequilibrium between markers in a single gene region. This procedure yielded a gene-level adjusted P value for each SNP. In addition, a Bonferroni correction for the number of gene regions investigated was applied to determine the α level for pathway-level significance (i.e., adjustment across all candidate gene regions; $0.05/53 = 9.43 \times 10^{-4}$ for adjusted P value). All analyses were performed using the statistical package R 2.81 (63).

Children's Health Study

Study sample and genotyping. The CHS is a longitudinal study designed to assess the long-term effects of air pollution

on children's health in the Los Angeles, California, area (64). The study consists of 5 cohorts totaling more than 11,000 children enrolled in 1993, 1996, and 2003, followed every year for up to 10 years. Genome-wide genotyping using the Illumina 550 K chip was recently performed on over 4,000 of the CHS subjects, with a primary focus of finding genes related to asthma and lung function development.

For this paper, we focused on the available data on the first wave of 1,942 children genotyped. We applied a stringent data-cleaning procedure that involves a quality control process of 30 or more steps, including identification and comparison of inter- and intraplate controls, stratified SNP and sample call rate investigations by cohort, confirmation of strand, concordance comparisons with additional genotyping performed via the Illumina GoldenGate Genotyping Assay, intensity checks, identification of chromosome X and Y anomalies, stratification by stripe, and tests of Hardy-Weinberg equilibrium. The quality control procedures that we followed have become the general standard for genome-wide association studies. Specifically for the SNPs replicated in the CHS, call rates were greater than 94.7%, and all were in Hardy-Weinberg equilibrium ($P > 0.20$).

Of the 1,942 genotyped subjects, information on smoking behavior was available for 926 (the other 1,016 subjects belong to the recently enrolled (2003) cohort and consist mostly of children under 10 years of age). To ensure genetic and environmental homogeneity within the CHS sample and to reduce the impact of potentially different linkage disequilibrium patterns across ethnic groups, we excluded 322 children who identified themselves as Hispanic or had an estimated European coefficient of ancestry less than 90% from the population structure analysis, using the program STRUCTURE (63). We further excluded 1 child who initiated smoking before enrollment in the CHS. Thus, 603 Caucasian children were included in the final analysis. All CHS subjects and their parents gave informed consent, and the study was approved by the University of Southern California Institutional Review Board.

During the investigation, CHS subjects were interviewed annually to estimate the number of cigarettes they had smoked in the previous year before each interview. The phenotype of smoking initiation was defined as the first time a participant reported having ever smoked during the past year.

Statistical analysis. To validate the positive findings from the WSPT, we performed a Cox regression analysis to estimate the hazard rate ratio for smoking initiation over the follow-up period. Gender, the coefficient of European ancestry, and an indicator of cohort (the subjects in the first wave of genotyping came from different cohorts) were included in the Cox regression model as covariates.

RESULTS

Descriptive analysis

Wuhan Smoking Prevention Trial. The demographic characteristics of the WSPT sample are shown in Table 1. Of the 2,339 persons included in the final analysis, 706 (30.2%) had initiated smoking and 1,633 (69.8%) were

never smokers at the time of interview. The sample contained slightly more males (1,218; 52.1%) than females (1,121; 47.9%), and the prevalence of smoking initiation was much higher in males (40.9%) than in females (18.6%), corresponding to a gender odds ratio of 3.08 (95% confidence interval (CI): 2.55, 3.72). The average age of the subjects was 12.6 years (standard deviation, 0.7).

Children's Health Study. The CHS sample contained 285 males (47.3%) and 318 females (52.7%) (Table 1). The children were enrolled in the CHS at a mean age of 10.2 years (standard deviation, 0.96), and the average follow-up time was 7.41 years (standard deviation, 1.07). During follow-up, 59 (20.7%) of the males and 70 (22.0%) of the females initiated smoking. No gender difference in the rate of smoking initiation ($P = 0.74$) was observed.

Association with smoking initiation in all participants

Web Figure 1 shows adjusted P values for the association between the SNPs in the dopamine pathway and smoking initiation in the WSPT. The complete results for this analysis are listed in Web Table 2.

Several SNPs in the corticotropin-releasing hormone receptor 1 gene (*CRHR1*) (rs4076452, rs12953076, and rs4074461) were associated with smoking initiation with marginal statistical significance in the WSPT. Details on the association between the 3 SNPs and smoking initiation are shown in Table 2. The 3 SNPs were in strong linkage disequilibrium with each other ($r^2 > 0.95$) and were therefore considered 1 signal. The most significant SNP, rs4076452 (minor allele frequency = 0.11; recessive heredity model; odds ratio = 4.47, 95% CI: 2.15, 9.31), was an imputed SNP with a very high imputation quality score (0.99).

The association between rs4076452 and smoking initiation was further examined in the CHS. No association was observed (hazard rate ratio = 1.49, 95% CI: 0.61, 3.68; P value = 0.38). Additionally, we did not find any statistically significant association between the other 2 SNPs (rs12953076 and rs4074461) and smoking initiation in the CHS.

Association between *CALY* and smoking initiation in females

Considering the heterogeneity of smoking prevalence in the Chinese sample, we stratified the sample by gender and repeated the analyses described above. The gender-specific adjusted P values are shown in Web Figure 2, and the complete results can be found in Web Table 3.

In the WSPT, 1 SNP in the calcyon neuron-specific vesicular protein gene (*CALY*) (rs2298122, minor allele frequency = 0.07) and several other SNPs in linkage disequilibrium (rs11101694 and rs4838721, both with r^2 values of approximately 0.88 with rs2298122) were found to be associated with smoking initiation only in females. These SNPs were also considered to be 1 signal. The prevalence of smoking initiation in females was 16.6% with the TT genotype of rs2298122, 34.9% with the GT genotype, and 50.0% with the GG genotype. Assuming an additive heredity model, the corresponding odds ratio was 2.21

(95% CI: 1.49, 3.27), with a crude P value of 8.39×10^{-5} (Table 3). After adjustment for all of the SNPs within *CALY* and determination of the heredity model, the adjusted P value was 4.10×10^{-4} , which was significant after adjustment for all genes tested. No association between rs2298122 and smoking initiation was observed in males (odds ratio = 0.88, 95% CI: 0.63, 1.21; adjusted $P = 0.68$).

We further examined the association of rs2298122 with smoking initiation in the CHS. As shown in Figure 1, the smoking initiation rate among females with the TT genotype was consistently lower than that among those with the GG/GT genotypes. Assuming an additive heredity model, the estimated hazard rate ratio for rs2298122 in females was 2.05 (95% CI: 1.27, 3.31; $P = 0.003$). Consistent with the WSPT, this association was not observed in males (hazard rate ratio = 1.32, 95% CI: 0.78, 2.23; $P = 0.304$).

DISCUSSION

In the current investigation, we first explored the effect of genetic variants in the dopamine pathway on smoking initiation in a large, cross-sectional sample of Chinese Han adolescents. After an extensive pathway-based candidate gene selection, a comprehensive SNP selection, genotyping, and imputation for untyped SNPs, 2,723 SNPs in 57 genes were examined for association with smoking initiation. Thereafter, SNPs surviving stringent multiple-testing correction in the Chinese population were examined in an independent Caucasian adolescent cohort sample in Los Angeles, California, a dramatically different sociocultural setting. When screening for the effects of the candidate genes in the WSPT, 1 SNP in *CALY* (rs2298122) was associated with smoking initiation only in females. In the CHS, highly consistent patterns were observed for rs2298122. This replication in a completely independent adolescent sample from a different population with a distinct design appears promising in implicating *CALY* in smoking initiation.

Previous investigations have demonstrated that *CALY* might be important in brain function and psychiatric disorders. *CALY* codes for calcyon neuron-specific vesicular protein (65), a novel brain-specific protein. Recent investigations revealed that the calcyon C terminal can stimulate the self-assembly of clathrin in a dose-dependent fashion and is therefore involved in clathrin-mediated endocytosis, which is crucial for efficient synaptic transmission and optimizing the levels of releasable pools of neurotransmitters (66) and is thereby involved in dopamine-related signaling and dopamine activity. Both dopamine-related signaling and dopamine activity have been implicated in various brain functions, such as motor control and cognitive processing (67, 68). Specifically, it has been shown that dopaminergic neurotransmission plays an important role in impulsive decision-making (45, 69), sensation-seeking (70), novelty-seeking (71), and extraversion (72), all of which have been identified as risk factors for reaction to smoking cues and smoking initiation (46, 73, 74).

The top-hit SNP, rs2298122, is located in intron 5 of *CALY*, and the role of this variant in human cognition and

Table 1. Demographic Characteristics of Participants in the Wuhan Smoking Prevention Trial (1998–1999) and the Children’s Health Study (1993–2009), China and California

Study and Characteristic	Gender						Total		
	Male			Female			No.	%	Mean (SD)
	No.	%	Mean (SD)	No.	%	Mean (SD)			
Wuhan Smoking Prevention Trial									
Smoking initiation									
Yes	498	40.9		208	18.6		706	30.2	
No	720	59.1		913	81.5		1,633	69.8	
Total	1,218	100.0		1,121	100.0		2,339	100.0	
Age ^a , years									
Yes			12.65 (0.67)			12.52 (0.61)			12.62 (0.66)
No			12.49 (0.66)			12.44 (0.70)			12.46 (0.68)
Total			12.55 (0.67)			12.45 (0.68)			12.51 (0.68)
Children’s Health Study									
Smoking initiation									
Yes	59	20.7		70	22.0		129	21.4	
No	226	79.6		248	78.0		474	78.6	
Total	285	100.0		318	100.0		603	100.0	
Age at initiation ^b , years									
			15.12 (2.20)			14.99 (1.99)			15.05 (2.09)

Abbreviation: SD, standard deviation.

^a Age at interview.

^b Age at which the participant first reported smoking in the past year.

behavior is not well understood. Within HapMap 3, 2 SNPs in the CEU population (rs7085530 and rs2105341) and 4 SNPs in the CHB population (rs11101694, rs2275723, rs2105341, and rs7085530) are in notable linkage disequilibrium with rs2298122 ($r^2 > 0.5$). However, none of these SNPs are nonsynonymous SNPs. Among the few association studies on *CALY*, Laurin et al. (75) performed a nuclear-family-based association study of attention deficit/hyperactivity disorder in Caucasian adolescents. Although their results are only suggestive, the association between rs2298122 and the impulsive dimension of attention

deficit/hyperactivity disorder was marginally significant ($P = 0.08$) (75). The molecular mechanism for the association between *CALY* and smoking initiation might reflect a potential effect via impulsivity, which has been shown to play a complex role in smoking behavior and the reaction to smoking cues (45, 69).

The association between *CALY* and smoking initiation was observed only in females. It is well known that smoking behavior differs between the genders, especially in some populations, such as the Chinese (76). Furthermore, a recent investigation in 32,359 pairs of native California twins

Table 2. Association Between 3 Single Nucleotide Polymorphisms in the Corticotropin-Releasing Hormone Receptor 1 Gene (*CRHR1*) and Smoking Initiation in the Wuhan Smoking Prevention Trial (1998–1999) and the Children’s Health Study (1993–2009), China and California

Study and SNP	SNP Type	Minor Allele Frequency	Model	Odds Ratio or HRR ^a	95% Confidence Interval	Crude P Value	Adjusted P Value ^b
Wuhan Smoking Prevention Trial							
rs4076452	Imputed	0.11	Recessive	4.47	2.15, 9.31	6.35×10^{-5}	8.18×10^{-4}
rs12953076	Genotyped	0.10	Recessive	4.72	2.16, 10.30	1.00×10^{-4}	1.14×10^{-3}
rs4074461	Genotyped	0.11	Recessive	3.91	1.94, 7.88	1.43×10^{-4}	2.37×10^{-3}
Children’s Health Study							
rs4076452	Genotyped	0.15	Recessive	1.49	0.61, 3.68	0.384	NA
rs12953076	Imputed	0.15	Recessive	1.45	0.59, 3.59	0.422	NA
rs4074461	Genotyped	0.47	Recessive	0.75	0.48, 1.17	0.198	NA

Abbreviations: HRR, hazard rate ratio; NA, not applicable; SNP, single nucleotide polymorphism.

^a Odds ratio (Wuhan Smoking Prevention Trial) or HRR (Children’s Health Study) estimate for the SNP with the putative heredity model.

^b P value after accounting for the multiple testing due to heredity model selection and multiple SNPs in the given gene region.

Table 3. Association Between the rs2298122 Single Nucleotide Polymorphism in the Calcyon Neuron-Specific Vesicular Protein Gene (*CALY*) and Initiation of Smoking, by Gender, Wuhan Smoking Prevention Trial (1998–1999) and the Children’s Health Study (1993–2009), China and California

Study and Gender	SNP Type	Minor Allele Frequency	Model	Odds Ratio or HRR ^a	95% Confidence Interval	Crude P Value	Adjusted P Value ^b
Wuhan Smoking Prevention Trial							
Male	Genotyped	0.07	Additive	0.88	0.63, 1.21	0.427	0.680
Female	Genotyped	0.06	Additive	2.21	1.49, 3.27	8.39×10^{-5}	4.10×10^{-4}
Total	Genotyped	0.07	Additive	1.30	1.00, 1.68	0.049	0.553
Children’s Health Study							
Male	Genotyped	0.18	Additive	1.32	0.78, 2.23	0.304	NA
Female	Genotyped	0.20	Additive	2.05	1.27, 3.31	0.003	NA
Total	Genotyped	0.19	Additive	1.69	1.19, 2.4	0.004	NA

Abbreviations: HRR, hazard rate ratio; NA, not applicable; SNP, single nucleotide polymorphism.

^a Odds ratio (Wuhan Smoking Prevention Trial) or HRR (Children’s Health Study) estimate for the SNP with the putative heredity model.

^b P value after accounting for the multiple testing due to heredity model selection and multiple SNPs in the given gene region.

indicated a gender difference in the pattern of genetic and environmental determinants of smoking initiation (77). Moreover, a gender difference in the association between some psychological characteristics and smoking-related behaviors has been repeatedly reported (78, 79). These studies indicate that there may be a subtle difference in the mechanisms or underlying genetic factors for smoking initiation between the genders. Our observation of an association only in females may partly reflect this.

In the WSPT, we also observed a strong effect of *CRHRI*, although this association was not replicated in the CHS. *CRHRI* codes for the type I receptor of corticotrophin-releasing hormone (80), and it has been associated with depression (81, 82), suicide (83), and alcohol dependence (84, 85) in previous investigations. However, the role of *CRHRI* in smoking-related behaviors has not been demonstrated. There are 2 possible explanations for the lack of replication of the positive findings in the CHS. First, in the WSPT, this association was only marginally significant for a pathway-wide significance level. Considering the fact that the best-fitting heredity model for this SNP was recessive, sparse data bias (86) might have arisen and led to unstable estimates or false positives. Furthermore, even if the findings for *CRHRI* in WSPT are true, statistical power to detect the recessive genetic effects would have been low in the CHS because of the limited sample size. Second, none of the *CRHRI* SNPs associated with smoking initiation in the WSPT are located in the coding region, indicating that those SNPs might be markers for another rare SNP in linkage disequilibrium with them in a Chinese population. Therefore, the different linkage disequilibrium structure between the Chinese and Caucasian populations might have led to inconsistent association results in the 2 populations. In either scenario, further functional or population-based data may help to clarify the role of *CRHRI* in smoking initiation.

The primary phenotype of the current investigation, smoking initiation in young adolescents, is different from lifetime smoking initiation. Specifically, given the high prevalence of smoking in adult Chinese males (>60%) (75), some persons

who will initiate smoking later in adolescence or adulthood might have been treated as controls in the WSPT. This may have led to decreased statistical power if we believe that lifetime smoking initiation is a homogenous phenotype. However, since smoking is integral to social interactions among adults in the Chinese population, there may exist different mechanisms or competing risk factors for smoking initiation at later ages, as compared with early smoking initiation, in which personality factors such as impulsivity, behavior control, decision-making, and novelty-seeking might play more important roles—especially in adolescent females. In addition, we limited our analysis in the WSPT to smoking status at the time of enrollment and did not include longitudinal follow-up data. This was primarily done to avoid any heterogeneity in inference that may have existed postintervention, since the WSPT was specifically designed to test the effectiveness of a school-based intervention aimed at altering adolescent smoking trajectories. In contrast, because the CHS is a cohort representing more natural smoking trajectories, our replication analysis utilized the entire longitudinal history. If we restrict our analysis to comparable ages (12 years and 13 years), we see that the pattern of association remains (odds ratio = 2.25, 95% CI: 0.9, 5.61), although it is not statistically significant because of decreased information in the small sample.

Subtle differences between early and lifetime initiation highlight the difficulty of research in this area, especially when considering the numerous nongenetic (host, social, and environmental) factors that are known to influence smoking initiation. The impact and temporal dynamics of these factors may lead to many gene-environment interactions, making discovery and replication across cohorts more difficult. As an example, in 2 recent genome-wide association studies on smoking behavior (Cancer Genetic Markers of Susceptibility and Genetic Association Information Network), no significant association (defined as $P < 0.00001$) between adult smoking initiation and variants in the *CALY* region (33, 34) was identified in an analysis combining both males and females. These differences in findings may be partly attributed to differences in the definition of smoking

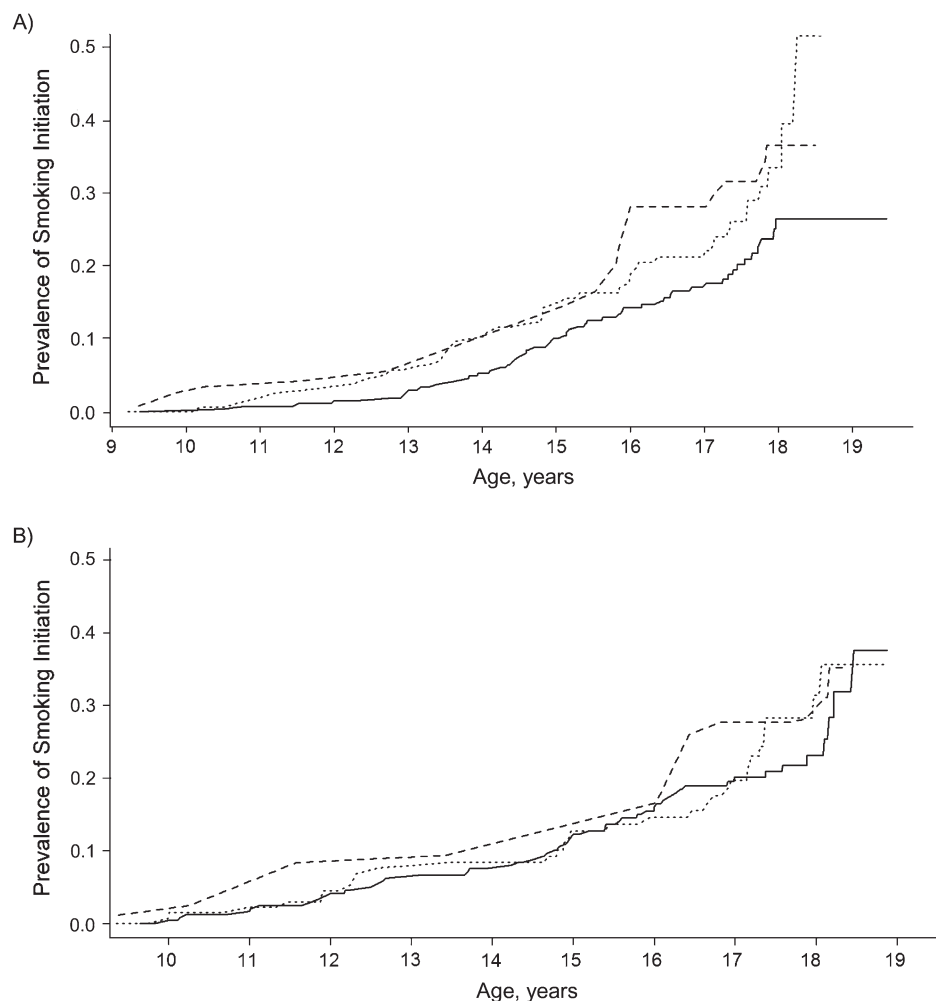


Figure 1. Association of the rs2298122 single nucleotide polymorphism in the calcyon neuron-specific vesicular protein gene (*CALY*) with smoking initiation in A) males and B) females during follow-up, Children's Health Study, California, 1993–2009. — —, GG genotype; - - -, TG genotype; —, TT genotype.

initiation, gender interaction, or differences in exposure to nongenetic factors. In any case, further replication and characterization via stratified analysis is required in this area of genetic research.

While a long-term goal of research in this area is aimed at elucidating potential interactions, we focused the current analyses on main effects of genes and heterogeneity due to gender. Although the positive findings of our exploratory investigation in the WSPT are supported by an independent sample, there are no functional data to further illustrate the putative association. Furthermore, when the WSPT cohort study was initiated in 1998, information on related psychological traits associated with smoking initiation, such as impulsivity and novelty-seeking, was not collected, and measures of depression and hostility were not derived from standardized, established items. This has greatly limited our ability to explore the role of personality characteristics as mediators of or in-

termediates in the association between the dopamine pathway and smoking initiation. Despite these caveats, we believe the evidence is sufficient to encourage further investigation to verify the potential role of *CALY* variants in smoking initiation.

In summary, we performed a pathway-based association study in a homogenous Chinese Han adolescent population to explore the role of the dopamine pathway in smoking initiation. We identified 1 SNP (rs2298122) in the *CALY* gene that was positively associated with smoking initiation only in females. Supportive evidence for this association was subsequently observed in an independent sample of Caucasian adolescents. While these findings will ultimately need to be replicated in additional population-based samples and explored in functional studies, the results of our investigation provide a foundation for future research in understanding the genetic mechanism of smoking initiation in adolescents.

ACKNOWLEDGMENTS

Author affiliations: Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California (Dalin Li, Jinghua Liu, Wonho Lee, Xuejuan Jiang, David Van Den Berg, Jim Gauderman, Frank Gilliland, Chih-Ping Chou, David V. Conti); National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina (Stephanie J. London); Center for Health Sciences, SRI International, Menlo Park, California (Andrew W. Bergen, Denise Nishita, Gary E. Swan, Nahid Waleh); School of Community and Global Health, Claremont Graduate University, Claremont, California (Peggy Gallaher, Jennifer B. Unger, C. Anderson Johnson); and Department of Pharmacology and Pharmaceutical Sciences, School of Pharmacy, University of Southern California, Los Angeles, California (Jean C. Shih).

This work was supported by the US National Institute on Drug Abuse (grants DA020830 and CA084735 to Drs. Dalin Li and David V. Conti) and the US National Institute of Environmental Health Sciences (grants ES015090 and GM069890 to Dr. David V. Conti). This work was also supported in part by the Intramural Research Program of the US National Institutes of Health, National Institute of Environment Health Sciences. Genotyping for the example was performed as part of the Pharmacogenetics of Nicotine Addiction Treatment Program (SRI International and University of California, San Francisco), which received funding from the National Institute on Drug Abuse (grant DA020830).

The authors acknowledge Drs. Yungang He, Ruhong Jiang, Li Su, Rich Flores, Suman Prasad, and Huijun Ring for their contributions to DNA analysis in the Wuhan Smoking Prevention Trial.

Conflict of interest: none declared.

REFERENCES

- Annual smoking-attributable mortality, years of potential life lost, and productivity losses—United States, 1997–2001. *MMWR Morb Mortal Wkly Rep.* 2005;54(25):625–628.
- World Health Organization. *Economics of Tobacco Control*. Geneva, Switzerland: World Health Organization; 2004. (www.who.int/gb/fctc/PDF/wg1/e1t2.pdf). (Accessed February 24, 2010).
- Warren CW, Jones NR, Peruga A, et al. Global youth tobacco surveillance, 2000–2007. *MMWR Surveill Summ.* 2008; 57(1):1–28.
- Peto R, Lopez AD, Boreham J, et al. *Mortality From Smoking in Developed Countries 1950–2000: Indirect Estimation From National Vital Statistics*. Oxford, United Kingdom: Oxford University Press; 2004.
- Chassin L, Presson CC, Rose JS, et al. The natural history of cigarette smoking from adolescence to adulthood: demographic predictors of continuity and change. *Health Psychol.* 1996;15(6):478–484.
- Everett SA, Warren CW, Sharp D, et al. Initiation of cigarette smoking and subsequent smoking behavior among U.S. high school students. *Prev Med.* 1999;29(5): 327–333.
- 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.
- Lando HA, Thai DT, Murray DM, et al. Age of initiation, smoking patterns, and risk in a population of working adults. *Prev Med.* 1999;29(6):590–598.
- Tyas SL, Pederson LL. Psychosocial factors related to adolescent smoking: a critical review of the literature. *Tob Control.* 1998;7(4):409–420.
- Unger JB, Cruz TB, Rohrbach LA, et al. English language use as a risk factor for smoking initiation among Hispanic and Asian American adolescents: evidence for mediation by tobacco-related beliefs and social norms. *Health Psychol.* 2000;19(5):403–410.
- Breslau N, Peterson EL, Schultz LR, et al. Major depression and stages of smoking. A longitudinal investigation. *Arch Gen Psychiatry.* 1998;55(2):161–166.
- Rodriguez D, Tercyak KP, Audrain-McGovern J. Effects of inattention and hyperactivity/impulsivity symptoms on development of nicotine dependence from mid adolescence to young adulthood. *J Pediatr Psychol.* 2007;33(6): 563–575.
- Bricker JB, Peterson AV, Robyn Andersen M, et al. Close friends', parents', and older siblings' smoking: reevaluating their influence on children's smoking. *Nicotine Tob Res.* 2006;8(2):217–226.
- Hoffman BR, Sussman S, Unger JB, et al. Peer influences on adolescent cigarette smoking: a theoretical review of the literature. *Subst Use Misuse.* 2006;41(1): 103–155.
- Bryant AL, Schulenberg J, Bachman JG, et al. Understanding the links among school misbehavior, academic achievement, and cigarette use: a national panel study of adolescents. *Prev Sci.* 2000;1(2):71–87.
- Hill KG, Hawkins JD, Catalano RF, et al. Family influences on the risk of daily smoking initiation. *J Adolesc Health.* 2005; 37(3):202–210.
- Unger JB, Baezconde-Garbanati L, Shakib S, et al. A cultural psychology approach to “drug abuse” prevention. *Subst Use Misuse.* 2004;39(10–12):1779–1820.
- Maes HH, Neale MC, Kendler KS, et al. Genetic and cultural transmission of smoking initiation: an extended twin kinship model. *Behav Genet.* 2006;36(6):795–808.
- Vink JM, Willemsen G, Boomsma DI. Heritability of smoking initiation and nicotine dependence. *Behav Genet.* 2005; 35(4):397–406.
- Schnoll RA, Johnson TA, Lerman C. Genetics and smoking behavior. *Curr Psychiatry Rep.* 2007;9(5):349–357.
- Li MD. The genetics of nicotine dependence. *Curr Psychiatry Rep.* 2006;8(2):158–164.
- David SP, Munafò MR. Genetic variation in the dopamine pathway and smoking cessation. *Pharmacogenomics.* 2008; 9(9):1307–1321.
- 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.
- Li MD. The genetics of smoking related behavior: a brief review. *Am J Med Sci.* 2003;326(4):168–173.
- Tyndale RF. Genetics of alcohol and tobacco use in humans. *Ann Med.* 2003;35(2):94–121.
- Morgan GD, Backinger CL, Leischow SJ. The future of tobacco-control research. *Cancer Epidemiol Biomarkers Prev.* 2007;16(6):1077–1080.

27. Li MD, Ma JZ, Beuten J. Progress in searching for susceptibility loci and genes for smoking-related behaviour. *Clin Genet*. 2004;66(5):382–392.
28. Chen X, Che Y, Zhang L, et al. *RhoA*, encoding a Rho GTPase, is associated with smoking initiation. *Genes Brain Behav*. 2007;6(8):689–697.
29. Rodríguez S, Huang S, Chen XH, et al. A study of *TH01* and *IGF2-INS-TH* haplotypes in relation to smoking initiation in three independent surveys. *Pharmacogenet Genomics*. 2006; 16(1):15–23.
30. Zhang L, Kendler KS, Chen X. Association of the phosphatase and tensin homolog gene (*PTEN*) with smoking initiation and nicotine dependence. *Am J Med Genet B Neuropsychiatr Genet*. 2006;141B(1):10–14.
31. David SP, Johnstone E, Griffiths SE, et al. No association between functional catechol *O*-methyl transferase 1947A>G polymorphism and smoking initiation, persistent smoking or smoking cessation. *Pharmacogenetics*. 2002;12(3):265–268.
32. Sullivan PF, Jiang Y, Neale MC, et al. Association of the tryptophan hydroxylase gene with smoking initiation but not progression to nicotine dependence. *Am J Med Genet*. 2001; 105(5):479–484.
33. Caporaso N, Gu F, Chatterjee N, et al. Genome-wide and candidate gene association study of cigarette smoking behaviors. *PLoS One*. 2009;4(2):e4653. (doi: 10.1371/journal.pone.0004653).
34. Vink JM, Smit AB, de Geus EJ, et al. Genome-wide association study of smoking initiation and current smoking. *Am J Hum Genet*. 2009;84(3):367–379.
35. Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genomewide association studies. *Am J Hum Genet*. 2007;81(6):1278–1283.
36. Conti DV, Cortessis V, Molitor J, et al. Bayesian modeling of complex metabolic pathways. *Hum Hered*. 2003;56(1–3): 83–93.
37. Janssens AC, van Duijn CM. Genome-based prediction of common diseases: advances and prospects. *Hum Mol Genet*. 2008;17(R2):R166–R173.
38. Phillips TJ, Belknap JK. Complex-trait genetics: emergence of multivariate strategies. *Nat Rev Neurosci*. 2002;3(6):478–485.
39. Gardner EL. Endocannabinoid signaling system and brain reward: emphasis on dopamine. *Pharmacol Biochem Behav*. 2005;81(2):263–284.
40. Feltstein MW, See RE. The neurocircuitry of addiction: an overview. *Br J Pharmacol*. 2008;154(2):261–274.
41. Dunlop BW, Nemeroff CB. The role of dopamine in the pathophysiology of depression. *Arch Gen Psychiatry*. 2007; 64(3):327–337.
42. Buddenberg TE, Topic B, Mahlberg ED, et al. Behavioral actions of intranasal application of dopamine: effects on forced swimming, elevated plus-maze and open field parameters. *Neuropsychobiology*. 2008;57(1–2):70–79.
43. Patton GC, Carlin JB, Coffey C, et al. Depression, anxiety, and smoking initiation: a prospective study over 3 years. *Am J Public Health*. 1998;88(10):1518–1522.
44. Althoa A, Seeman P, Kõiv K, et al. Rats with persistently high exploratory activity have both higher extracellular dopamine levels and higher proportion of D₂^{High} receptors in the striatum. *Synapse*. 2009;63(5):443–446.
45. Adriani W, Boyer F, Gioiosa L, et al. Increased impulsive behavior and risk proneness following lentivirus-mediated dopamine transporter over-expression in rats' nucleus accumbens. *Neuroscience*. 2009;159(1):47–58.
46. Audrain-McGovern J, Rodriguez D, Patel V, et al. How do psychological factors influence adolescent smoking progression? The evidence for indirect effects through tobacco advertising receptivity. *Pediatrics*. 2006;117(4):1216–1225.
47. Doran N, McChargue D, Spring B. Effect of impulsivity on cardiovascular and subjective reactivity to smoking cues. *Addict Behav*. 2008;33(1):167–172.
48. Chou CP, Li Y, Unger JB, et al. A randomized intervention of smoking for adolescents in urban Wuhan, China. *Prev Med*. 2006;42(4):280–285.
49. London SJ, Xia J, Lehman TA, et al. Collection of buccal cell DNA in seventh-grade children using water and a toothbrush. *Cancer Epidemiol Biomarkers Prev*. 2001;10(11): 1227–1230.
50. Warren CW, Riley L, Asma S, et al. Tobacco use by youth: a surveillance report from the Global Youth Tobacco Survey project. *Bull World Health Organ*. 2000;78(7):868–876.
51. Conti DV, Lee W, Li D, et al. Nicotinic acetylcholine receptor $\beta 2$ subunit gene implicated in a systems-based candidate gene study of smoking cessation. Pharmacogenetics of Nicotine Addiction and Treatment Consortium. *Hum Mol Genet*. 2008; 17(18):2834–2848.
52. Thomas PD, Mi H, Swan GE, et al. A systems biology network model for genetic association studies of nicotine addiction and treatment. *Pharmacogenet Genomics*. 2009; 19(7):538–551.
53. Meyer EL, Yoshikami D, McIntosh JM. The neuronal nicotinic acetylcholine receptors alpha 4* and alpha 6* differentially modulate dopamine release in mouse striatal slices. *J Neurochem*. 2008;105(5):1761–1769.
54. Edlund CK, Lee WH, Li D, et al. Snagger: a user-friendly program for incorporating additional information for tagSNP selection. *BMC Bioinformatics*. 2008;9(3):174. (doi: 10.1186/1471-2105-9-174).
55. Thorisson GA, Smith AV, Krishnan L, et al. The International HapMap Project Web site. *Genome Res*. 2005;15(11): 1592–1593.
56. Smith MW, Patterson N, Lautenberger JA, et al. A high-density admixture map for disease gene discovery in African Americans. *Am J Hum Genet*. 2004;74(5):1001–1013.
57. Shtir CJ, Marjoram P, Azen S, et al. Variation in genetic admixture and population structure among Latinos: the Los Angeles Latino Eye Study (LALES). *BMC Genet*. 2009; 10(10):71. (doi: 10.1186/1471-2156-10-71).
58. Deckert J, Catalano M, Sygailo, et al. Excess of high activity monoamine oxidase A gene promoter alleles in female patients with panic disorder. *Hum Mol Genet*. 1999; 8(4):621–624.
59. Li Y, Abecasis GR. Mach 1.0: rapid haplotype reconstruction and missing genotype inference. *Am J Hum Genet*. 2006; 79(suppl):2290.
60. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000; 155(2):945–959.
61. Gu D, Wu X, Reynolds K, et al. Cigarette smoking and exposure to environmental tobacco smoke in China: the International Collaborative Study of Cardiovascular Disease in Asia. *Am J Public Health*. 2004;94(11):1972–1976.
62. Conneely KN, Boehnke M. So many correlated tests, so little time! Rapid adjustment of *P* values for multiple correlated tests. *Am J Hum Genet*. 2007;81(6):1158–1168.
63. R Development Core Team. *R: a Language and Environment for Statistical Computing. Release 2.81*. Vienna, Austria: R Foundation for Statistical Computing; 2009.
64. Salam MT, Li YF, Langholz B, et al. Early-life environmental risk factors for asthma: findings from the Children's Health Study. *Environ Health Perspect*. 2004;112(6):760–765.

65. Lezcano N, Mrzljak L, Eubanks S, et al. Dual signaling regulated by calcyon, a D1 dopamine receptor interacting protein. *Science*. 2000;287(5458):1660–1664.
66. Xiao J, Dai R, Negyessy L, et al. Calcyon, a novel partner of clathrin light chain, stimulates clathrin-mediated endocytosis. *J Biol Chem*. 2006;281(22):15182–15193.
67. Crocker AD. The regulation of motor control: an evaluation of the role of dopamine receptors in the substantia nigra. *Rev Neurosci*. 1997;8(1):55–76.
68. Seamans JK, Yang CR. The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Prog Neurobiol*. 2004;74(1):1–58.
69. van Gaalen MM, van Koten R, Schoffelmeer AN, et al. Critical involvement of dopaminergic neurotransmission in impulsive decision making. *Biol Psychiatry*. 2006;60(1):66–73.
70. Netter P, Hennig J, Roed IS. Serotonin and dopamine as mediators of sensation seeking behavior. *Neuropsychobiology*. 1996;34(3):155–165.
71. Zald DH, Cowan RL, Riccardi P, et al. Midbrain dopamine receptor availability is inversely associated with novelty-seeking traits in humans. *J Neurosci*. 2008;28(53):14372–14378.
72. Depue RA, Collins PF. Neurobiology of the structure of personality: dopamine, facilitation of incentive motivation, and extraversion. *Behav Brain Sci*. 1999;22(3):491–517.
73. Lipkus IM, Barefoot JC, Feaganes J, et al. A short MMPI scale to identify people likely to begin smoking. *J Pers Assess*. 1994;62(2):213–222.
74. Van Loon AJ, Tjhuis M, Surtees PG, et al. Determinants of smoking status: cross-sectional data on smoking initiation and cessation. *Eur J Public Health*. 2005;15(3):256–261.
75. Laurin N, Misener VL, Crosbie J, et al. Association of the calcyon gene (*DRD1IP*) with attention deficit/hyperactivity disorder. *Mol Psychiatry*. 2005;10(12):1117–1125.
76. Yang T, Li F, Yang X, et al. Smoking patterns and sociodemographic factors associated with tobacco use among Chinese rural male residents: a descriptive analysis. *BMC Public Health*. 2008;8(4):248. (doi: 10.1186/1471-2458-8-248).
77. Hamilton AS, Lessov-Schlaggar CN, Cockburn MG, et al. Gender differences in determinants of smoking initiation and persistence in California twins. *Cancer Epidemiol Biomarkers Prev*. 2006;15(6):1189–1197.
78. Larsen JK, Otten R, Engels RC. Adolescent depressive symptoms and smoking behavior: the gender-specific role of weight concern and dieting. *J Psychosom Res*. 2009;66(4):305–308.
79. Fields S, Collins C, Leraas K, et al. Dimensions of impulsive behavior in adolescent smokers and nonsmokers. *Exp Clin Psychopharmacol*. 2009;17(5):302–311.
80. Timpl P, Spanagel R, Sillaber I, et al. Impaired stress response and reduced anxiety in mice lacking a functional corticotropin-releasing hormone receptor 1. *Nat Genet*. 1998;19(2):162–166.
81. Utge S, Soronen P, Partonen T, et al. A population-based association study of candidate genes for depression and sleep disturbance. *Am J Med Genet B Neuropsychiatr Genet*. 2009;153B(2):468–476.
82. Papiol S, Arias B, Gastó C, et al. Genetic variability at HPA axis in major depression and clinical response to antidepressant treatment. *J Affect Disord*. 2007;104(1–3):83–90.
83. Wasserman D, Sokolowski M, Rozanov V, et al. The *CRHR1* gene: a marker for suicidality in depressed males exposed to low stress. *Genes Brain Behav*. 2008;7(1):14–19.
84. Hansson AC, Cippitelli A, Sommer WH, et al. Variation at the rat *Crhr1* locus and sensitivity to relapse into alcohol seeking induced by environmental stress. *Proc Natl Acad Sci U S A*. 2006;103(41):15236–15241.
85. Treutlein J, Kissling C, Frank J, et al. Genetic association of the human corticotropin releasing hormone receptor 1 (*CRHR1*) with binge drinking and alcohol intake patterns in two independent samples. *Mol Psychiatry*. 2006;11(6):594–602.
86. Greenland S, Schwartzbaum JA, Finkle WD. Problems due to small samples and sparse data in conditional logistic regression analysis. *Am J Epidemiol*. 2000;151(5):531–539.