Environmental and Genetic Determinants of Tobacco Use: Methodology for a Multidisciplinary, Longitudinal Family-Based Investigation


Abstract
This article describes the ongoing collaborative effort of six research teams to operationalize and execute an integrative approach to the study of gene × environment interactions in the development of tobacco dependence. At the core of the project is a longitudinal investigation of social and behavioral risk factors for tobacco use in individuals who were, on average, 13 years of age at intake and for whom smoking outcomes extending from early adolescence to young adulthood have been characterized previously (current average age of the cohort is 29 years). The conceptual framework for the integrative approach and the longitudinal investigation on which the study is based is presented. A description is also provided of the methods used to: (a) recruit participants and families to provide DNA samples and information on tobacco use; (b) assess participants for relevant tobacco-related phenotypes including smoking history, current use of tobacco, and nicotine metabolism; (c) assess the quality of the DNA samples collected from participants for genome-wide scanning and candidate gene analysis; (d) examine several research questions concerning the role of genetic and environmental factors in the onset and maintenance of tobacco use; and (e) ensure adherence to local and federal guidelines for ethical and legal investigations of genotypic associations with tobacco-related phenotypes in families. This investigation is unique among ongoing studies of the genetics of tobacco dependence in the extent to which equal importance has been assigned to both phenotypic and genotypic measurements.

Introduction
Tobacco dependence and associated difficulty with quitting are hypothesized to result from a plethora of environmental factors and pharmacological effects of nicotine (1), a substance that is now widely recognized to be highly addictive (2). Environmental influences on the development of tobacco dependence are well documented, and include peer and familial influences as the strongest contributors in determining how and when cigarette experimentation occurs among young people (reviewed in Refs. 3–5). However, a largely independent literature of studies based on twin samples supports the hypothesis that genetic influences also underlie the initiation and lifetime use of tobacco (6–8).

A number of studies have examined the association between gene variants in neurotransmitter systems that are thought to play a role in the development of tobacco dependence. For example, associations between various components of the dopaminergic (i.e., DRD2 and SLC6A3; Refs. 9–11), opioidergic (i.e., OPRM1; Refs. 12, 13), nicotinic receptors (i.e., CHRNA4 and CHRNA5; Refs. 14–16), and metabolic pathways (i.e., CYP2D6 and CYP2A6; Refs. 17–19) with various smoking behaviors have been reported. Reports of specific genomic regions implicated as markers of susceptibility for smoking have also been reported in the scientific literature (20). Many of these studies have reported associations that, while statistically significant, are quite small in absolute terms and have not yet been replicated (compare 21).

Using advances made in the field of cancer epidemiology with regard to gene × environment interactions as a guide (22–25), it is reasonable to hypothesize that tobacco dependence also might be influenced by gene × environment interactions (26–28). Tobacco dependence, similar to certain cancers, exhibits characteristics of a complex genetic trait (e.g., gene × gene and gene × environment interactions, and phenotypic and genotypic heterogeneity; Refs. 29–31). It has been shown that, depending on the nature of the gene × environment interaction (e.g., additive, multiplicative, or synergistic), the potential impact on prevention of regular tobacco use could be quite high, thereby providing substantial motivation for the pursuit of an integrative understanding of how genes and the environment interact to increase susceptibility for tobacco use (28).

The purpose of this report is to describe a research ap-
An integrative model is proposed that integrates methods from genetic epidemiology, tobacco dependence assessment, nicotine pharmacology, and molecular genetic methods with a pre-existing study of the social and behavioral components of tobacco use and dependence. In this study, a series of hypotheses will be tested: (a) phenotypic definitions of tobacco dependence will vary in the extent to which they demonstrate familial aggregation and/or genetic association; (b) phenotypes that characterize a developmental trajectory of dependence behaviors (e.g., time from experimentation to regular use) or underlying nicotine biology (e.g., nicotine metabolism) will demonstrate stronger evidence of genetic involvement than will phenotypes that rely on broad nonspecific self-reports of smoking behavior (e.g., “ever-smoking”) or global measures of dependent tobacco use (e.g., the Fagerström Test for Nicotine Dependence, FTND; 32); (c) composite phenotypes, consisting of combinations of phenotypes with the strongest association with tobacco dependence, will demonstrate the strongest evidence of genetic influences; (d) genetic markers associated previously with tobacco-related phenotypes in case-control studies will be confirmed by genetic association designs that take into account genetic stratification; (e) the presence of environmental risk factors will mediate the association between candidate genes and tobacco dependence; (f) the genetic factors identified will be associated with longitudinal phenotypes (e.g., rate of progression toward regular smoking) identified by latent growth curve modeling; and, (g) a combination of genetic and environmental risk factors will differentiate participants who experienced early versus late onset of smoking.

An Integrative Model. The integrative model of tobacco use and dependence used to guide the efforts of this study (Fig. 1) recognizes the role played by individual differences in vulnerability factors (33), in tobacco use trajectories (34–36), environmental exposure (37), and in nicotine metabolism and dependence, including motivations to smoke and the reinforcement derived from tobacco (38, 39). Certain factors such as anxiety, depression, use of other substances, and family history of tobacco use, along with biologically-based individual differences in sensitivity, metabolism, and/or response to nicotine, might themselves have genetic components (33). It also has been suggested that the effects of these variables on subsequent likelihood of smoking are mediated by personal factors such as lower intelligence, socioeconomic status, and the occurrence of events within the social environment, such as peers who smoke and family discord (e.g., Refs. 33, 40).

Nicotine, the psychoactive alkaloid found in tobacco products, is thought to play a major role in tobacco dependence. Most smokers tend to ingest similar amounts of nicotine from day to day, consistent with the idea that they titrate their dose of nicotine to achieve desired effects (41). Nicotine is extensively metabolized in the body, primarily by the liver cytochrome P450 enzyme CYP2A6 (42, 43). Whereas previous research suggests a relationship between the CYP2A6 genotype and the likelihood of nicotine addiction (19), sample sizes in these studies are small, and findings are inconclusive. Nonetheless, because CYP2A6 activity affects the rate at which nicotine is eliminated, genetic alterations in the CYP2A6 enzyme may impact smoking behavior and dependence, and deserve additional attention. Other genetic factors that may contribute to tobacco dependence include variation in pathways responsible for nicotine reward and pleasure (44–46). An important feature of the model for tobacco use over the life span is that genetic and environmental factors exerting influence at different points in the development of tobacco use (e.g., initiation, maintenance, cessation, and relapse) are likely to be different (47). The integrative model described herein considers
many of these genetic and environmental factors described above in the initiation, establishment, and maintenance of tobacco use and dependence.

Materials and Methods

The IRP: An Overview

This IRP, funded by the University of California Tobacco-Related Disease Research Program, uses data from the SMOFAM study (DA03706, Hyman Hops, PI, Oregon Research Institute). The SMOFAM study is a comprehensive, repeated measures cohort study of environmental and psychosocial risk factors for adolescent and young adult substance use, including tobacco. The original SMOFAM study, initiated in 1984, recruited 763 adolescents (referred to throughout this article as “target” participants) 11–15 years of age, their parents, and siblings for a 3-year longitudinal study. Over a 17-year period, 15 annual assessments have been completed and as of the last assessment, there were 465 target participants ranging from 27 to 31 years of age remaining in the study. The repeated assessment of target participants facilitated characterization of longitudinal phenotypes for tobacco use, including the acquisition and maintenance of smoking, as well as many potential psychosocial and environmental predictors of substance use including peer and parental approval of smoking (48), parenting behavioral problems (49), aggressive behavior (50), depressive symptoms (51), conflict behavior (52), and family environment (53). Since its inception, the SMOFAM study has been the source of >30 scientific publications (e.g., Refs. 54–57).

As shown in Fig. 2, the IRP consists of three independent but interrelated studies coordinated by a statistical core. Study 1 has reorganized the existing longitudinal SMOFAM data into a pedigree structure and will provide an analysis of tobacco use patterns in families. The main purpose for study 2 is the identification of genetic variants associated with tobacco-related phenotypes using both a candidate gene approach as well as a genome-wide scan to test for loci undetected previously. Investigators involved in study 2 work closely with the statistical core to conduct genetic association and linkage analyses. Finally, the main purpose of study 3 is to examine familial aggregation of nicotine and cotinine metabolic phenotypes. Together, these study components represent a unique collaboration and ability to apply a comprehensive approach for in-
vestigating factors associated with the initiation, establishment, and maintenance of tobacco dependence in families. An important additional component is the systematic consideration of bioethical implications of the project (Stanford Center for Biomedical Ethics). Below, the methodology underlying the IRP and the manner in which the individual components interact is described (see also Fig. 3 for an illustration of the flow of communication and study data between components). Communication between the study centers occurs primarily through e-mail correspondence and telephone, with monthly team conference calls. Subcommittees of study investigators meet periodically to address key issues, such as to develop study measures and to delineate analytic approaches. Annually throughout the course of the study, all of the investigators and support staff convene for a 2-day working group meeting.

**IRP Study 1: Analysis of Tobacco Use Patterns in Families**

The primary goal of study 1 is to determine whether patterns of tobacco use among first-degree biological relatives demonstrate evidence of familial aggregation and, if so, whether the observed familial clustering suggests the presence of genetic and/or environmental transmission from parents to offspring (compare Ref. 58). Telephone interviews with the target participants to assess family history of tobacco use were conducted, and written smoking history questionnaires were administered to targets and their biological relatives to additionally characterize tobacco-related phenotypes. These two study instruments, administered in the initial year of our study, are described below.

**FHTU Interview to Identify Prospective IRP Participants.** Targets who completed at least 7 of the first 10 assessments in the SMOFAM study were recruited for the FHTU interview, and a total of 481 (92.5%) completed it. Specific variables that were assessed in the FHTU interview for each first degree relative include: sex, age, relationship (biological or nonbiological; full-, half-, or nonbiological sibling), vital status, lifetime “ever” smoking of ≥100 cigarettes, ever regular use of cigars, pipes, or smokeless tobacco, age at initiation of daily smoking, and age at cessation of smoking.

**Study 1 administers FHTU to SMOFAM targets.**

**Core prioritizes recruitment efforts and administers SHQ to targets and family members.**

**Core collects blood samples from individuals within families who meet the minimum criteria for participation.**

**Using SHQ data, Study 1 defines tobacco related phenotypes, conducts segregation and heritability analyses.**

**Core delivers blood samples to Study 2 and provides subject referrals to Study 3’s component.**

**Study 2 genotypes genome-wide marker set and candidate genes.**

**Study 3 administers oral nicotine administration protocol, collecting saliva and urine specimens.**

**Core and Study 2 independently perform linkage and association analyses.**

**Study 3 completes assays; passes nicotine metabolism data to Core and Study 2.**

**All IRP components review results of linkage and association analyses.**

**Fig. 3.** Flow of data between components of the IRP.
cigarette smoking, average number of cigarettes smoked per day when smoking, ever tried to quit, and success in permanent quitting. The objectives of the FHTU interview were 4-fold: (a) to confirm the existence of biological relationships among the target, family members, and other household members who participated in SMOFAM; (b) to identify the additional first-degree relatives of the target who did not participate in SMOFAM; (c) to identify other nonbiological family members, spouses, or partners with whom the target had lived, regardless of their SMOFAM participation; and (d) to collect the report by the target of the tobacco use history of all identified individuals. On average, the FHTU took 15 min to complete.

Prioritizing Families for Recruitment into the Full Study. On the basis of data from the FHTU, families were stratified by the reported level of “density” of the ever-smoking phenotype (e.g., the number of biological members who had smoked >100 cigarettes in lifetime) within the nuclear family. An essential criterion for prioritizing families was that the family should be composed of at least two living, ever-smoking members who were fully biologically related to the target. The target could be counted as one of the two ever-smoking members. In anticipation of the measured genetic component of the planned analysis (see study 2 below), eligible families must have consisted of a minimum “triad” configuration, defined as two biological parents plus an ever-smoking offspring, or a biological sibling pair plus a biological parent (with one ever-smoker among them). Regardless of their individual ever-smoking status, the target and all of the first-degree relatives within these families were invited to complete the SHQ and to provide a blood sample. Targets from families not meeting the minimum triad configuration were asked to complete the SHQ but not blood collection.

SHQ. A comprehensive tool for assessing past and current tobacco use was developed to validate and supplement existing SMOFAM data, as well as the data provided by the target via the FHTU interview. The 24-page instrument took ~30 min to complete and assessed the following: (a) sociodemographics; (b) general health; (c) smoking environment, including number of hours of environmental tobacco smoke exposure in the workplace and home each day; (d) alcohol intake; (e) use of cigars, pipes, smokeless tobacco, and marijuana; (f) early experiences with cigarette smoking (59); (g) cigarette smoking history; (h) daily smoking experiences (60); (i) nicotine dependence and withdrawal as assessed by the Fagerström Tolerance Questionnaire/Fagerström Test for Nicotine Dependence (32, 61) and the Nicotine Dependence Scale (62); (j) SMQ (63); (k) Nicotine Reinforcement Questionnaire (64); (l) quitting and withdrawal history; (m) social support for quitting; and (n) depressive symptoms as assessed by the Centers for Epidemiological Studies 20-item scale, for use in the general population (the CES-D; 51).

Measures of Other Tobacco-Related Phenotypes. A consensus has emerged in which tobacco dependence is viewed as multidimensional and should be assessed and quantified according to a wide array of tobacco-related phenotypes was assessed for all of the participants via the FHTU and SHQ tools (hard copies of each are available from the senior author on request). Also, because in utero exposure to nicotine may be a risk factor for later smoking (70), the biological mother of the target completed an additional survey of her own tobacco, alcohol, and other drug use during each pregnancy she carried to term.

IRP Study 2: Identification of Genetic Variants Associated with Tobacco Use

Study 2 is ongoing and involves two different methods, the genome-wide scan and the candidate gene approaches. DNA was isolated from frozen blood samples using a standard commercial DNA isolation kit (Puregene, Gentra Systems, Inc., Minneapolis, MN). DNA concentration was determined by 260/280 absorbance. The DNA produced by this method has an average size >20 kbp; all of the samples were determined to have adequate amplification in response to test PCR reactions.

Genome-Wide Scan and Quality Assurance. A genome-wide scan was undertaken to confirm genetic linkages reported previously and/or to identify new genomic regions of interest. DNA from all of the available family members was genotyped for an 811 dinucleotide repeat microsatellite marker (71) panel (HDS; Applied Biosystems, Foster City, CA). The markers are distributed throughout the genome with a marker-to-marker distance of <5 cm, a smaller distance than the 10-cm interval used in the five linkage studies of tobacco use published to date (20, 72–75).

Accuracy of pedigree structure and genotype determination is crucial for the planned genetic analyses, because it has been suggested that as small as a 1% genotype error rate can significantly affect the analysis of a genomic linkage scan (76). A multistage data analysis approach was used to minimize errors in pedigree structure, sample identity, and genotypes. Genotypes for this genomic scan were used to confirm sample identity and pedigree structure by estimation of the average allele sharing between participants (77). Participant pairs that did not have the expected allele sharing were regenotyped using DNA from the reserve blood specimen. Genotypes that appeared to be erroneous based on comparison to genotypes of linked markers (78) were reviewed, edited, or suppressed. The high density of the marker panel and the structure of the pedigrees indicated there was little loss of segregation information. The incidence of genotypes determined blind to pedigree structure that were inconsistent with Mendelian segregation in a pilot study of 50 nuclear families was 0.2%. Review and replication of genotypes that were inconsistent with Mendelian segregation errors clearly indicated that many of these genotypes were reproducible and probably represented somatic mutations that would have interfered with analysis if not recognized.

Candidate Gene Approach. As described previously, DNA from study participants will be genotyped for variants of genes of interest (see “Introduction”). Allelic association analysis will be used to detect the effects of gene variants on smoking behavior. Using this approach, efforts will be focused on detecting common variants (e.g., prevalence >5%) that predispose an individual to smoking behaviors. Common variants are likely to be in linkage disequilibrium with haplotypes of other tightly linked common variants. Therefore, an average of 10 single nucleotide polymorphisms (SNP) for a gene of average size (~40 kbp) will be investigated. Polymorphisms will be prioritized based on the likelihood that they are informative and have a functional effect on the gene (missense changes and changes in evolutionarily conserved sequences are genotyped preferentially). Should a haplotype be associated with smoking behavior, a systematic search will be made for other polymorphisms that could have a functional effect on smoking behavior.
IRP Study 3: Assessment of Nicotine Metabolism Parameters in Nuclear Families

Study 3 is ongoing and is assessing the relationship between genetic factors and nicotine metabolism in a subsample of the SMOFAM families. In this study, nicotine and metabolite levels in saliva and urine are measured after administering oral doses of nicotine and its major metabolite, cotinine. Nicotine and cotinine are metabolized primarily by the liver enzyme CYP2A6 to cotinine and 3HC, respectively (42, 43). By measuring levels of cotinine in the saliva over time after an oral dose of cotinine, one can compute its oral clearance, which is a phenotypic marker of nicotine clearance (79). The ratio of 3HC:cotinine in the saliva or urine can also provide a measure of rate of metabolism of nicotine (80). This ratio reflects the rapidity of metabolism of cotinine to 3HC, which, in turn, reflects CYP2A6 activity (the higher the ratio, the greater the CYP2A6 activity). This ratio also has been shown to be highly correlated with the oral clearance of nicotine, confirming that the ratio is a marker for the speed of nicotine metabolism (81).

Finally, nicotine excretion and the ratio of 3HC:cotinine in the urine over 8 h will be measured, both of which also can be used as estimates of nicotine clearance.

Agent Administration and Sample Collection Procedures. Study 3 is carried out by a research nurse in the home of the participants. At the beginning of the study, fasting participants provide saliva and urine specimens, and a pregnancy test is performed on all of the females. A positive pregnancy test would end participation in study 3.

After oral administration of 1000 mg of ammonium chloride to acidify the urine, non-nicotine users are given unlabeled cotinine (cotinine-d0). Current nicotine users are given deuterium-labeled nicotine (3’,3’-dideuterocotinine; nicotine-d2) and deuterium-labeled cotinine (2,4,5,6-tetradecotinine; cotinine-d4), which are metabolized in the same way as are natural nicotine and cotinine, respectively (82). The advantage of deuterium-labeled compounds is that they can be measured by mass spectrometry even in smokers who have natural nicotine and cotinine present in their bodies from tobacco use.

Saliva samples, 5 ml each, are collected at 6, 12, 24, 36, 48, and 60 h. Urine is collected over a period of 8 h subsequent to oral dosing with nicotine/cotinine. The time, urine volume, and pH are recorded.

Aliquots of urine and saliva samples are frozen and shipped on dry ice in batches to the laboratory. Concentrations of cotinine and saliva are measured over time by gas chromatography-mass spectrometry (82), and the half-life and area under the saliva concentration time curve are determined. Cotinine oral clearance is computed as dose/area under the saliva concentration time curve. The 3HC:cotinine ratio is measured in saliva at 8 h. Urine is assayed for nicotine, cotinine, and 3HC, as well as other metabolites (83).

Ethical and Legal Considerations

Local, State, and Federal Regulations Pertaining to Informed Consent. Informed consent was obtained from all of the participants for the questionnaire components of the study. For the genotyping component of this study, informed consent was obtained from each participant who provided a blood sample. Children under the age of 18 were precluded from participation in the genotyping study. Explicit informed consent was also obtained from all of the participants in the nicotine and cotinine metabolism study (study 3). Along with approval from the Institutional Review Board for the Protection of Human Subjects from each of the collaborating institutions, study procedures were required to be in compliance with state laws dealing with genetics. To assure the maximum protection of participants in this study, a certification of confidentiality was obtained from the United States Department of Health and Human Services. Once a certificate was issued, only the Department of Health and Human Services could have access to information about participants for purposes of audit.

Informed Consent of Secondary Research Participants. The issue of who should be considered a research participant when participants are asked to provide information about other family members has received much attention recently (84). The other family members for whom information is provided by a research participant are sometimes referred to as “secondary participants.” The issue is whether such individuals meet the federal definition of a human subject and whether specific informed consent must be obtained from them. To create explicit protocols for determining what information needed to be obtained from research participants about secondary participants, whether explicit informed consent needed to be obtained from secondary participants, and how to protect their confidentiality to the greatest extent possible, the procedures available online (85) were followed.

State and Federal Guidelines and Interstate Issues for the Acquisition and Preservation of DNA Samples. In acquiring DNA samples, legal requirements specific to a particular state, as well as ethical principles defined by the Office for Protection from Research Risks at the federal level needed to be met. As a multistate study, the project was faced with the potentially time-consuming task of monitoring and abiding by an assortment of state rules given an absence of a single comprehensive federal law (86). In the present study, most participants who provided blood for genotyping studies were recruited from the state of Oregon, which has specific and particularly stringent legal requirements governing genetics. Other states, where study participants were recruited, specifically California and Washington, do not have legal requirements. Thus, the consent form was written to conform to Oregon state law on genetic privacy and used for all of the participants regardless of their state of origin. To meet the Oregon standards, the consent form specifically highlighted three points regarding the use of DNA samples for study participants: (a) researcher intention to retain the sample for additional studies; (b) the scope of the intended subsequent use; and (c) the right of the participants to have their sample destroyed.

Provision of Genetic Test Results to Participants. Because research to date has failed to establish a replicable linkage between a polymorphism and a specific risk for nicotine addiction, it was decided that the results of the genetic analyses would not be made available to study participants; this was disclosed in the informed consent process. Although it is possible that this genetic information might acquire value in the future, it was determined that the risk of disclosure of information that has no meaning today exceeds any perceived benefit. An additional consideration in the decision to not provide results of genetic tests to participants was derived from federal regulations. Specifically, the Clinical Laboratory Improvement Act (87) prohibits the disclosure of research results that are not obtained in a laboratory certified by the appropriate agencies overseeing laboratory quality. Such results, not validated for clinical use, may not be provided to participants.

1 Internet address: http://ohrp.osophs.dhhs.gov/ohrparc/mitg01–02/third.pdf.
Results

**FHTU Interviews and SHQ: Characterization of Nonresponders**

Four hundred eighty one targets completed the FHTU, providing information on 1,446 biological family members including 525 (35.3%) who had not participated in the original SMOFAM study. When compared with the targets from the original SMOFAM cohort who did not participate in the FHTU interview (n = 282), the targets who did were more likely to be female (55.1% versus 44.0%; \( \chi^2 = 8.8; P = 0.003 \)), and at the initial SMOFAM assessment were younger (13.1 versus 13.4 years; \( t (761) = 2.83; P < 0.005 \)), and less likely to report having smoked more than once (28.1% versus 41.7%; \( \chi^2 = 14.7; P < 0.0001 \)) and to report having fathers who ever smoked (60.9% versus 69.4%; \( \chi^2 = 4.8; P < 0.03 \)).

A total of 378 targets completed the SHQ (78.6% of the 481 targets completing the FHTU interview). As compared with the 103 targets that did not complete a SHQ, the 378 targets who did so had similar rates of ever-smoking (53.4% versus 41.7%; \( P = 0.003 \)), and at the initial SMOFAM assessment were younger (13.2; \( t (235.3) = 2.83; P < 0.005 \)) and had participated in more previous SMOFAM assessments (average of 9.0 versus 7.9 assessments [of 10 maximum]; \( t (134) = -3.76; P = 0.0003 \)).

**Recruitment of High Priority Families.** Two hundred thirty seven families (979 family members) were identified as “high priority” for recruitment according to the criteria described previously. As would be expected, when compared with the targets from the 244 “low-priority” families, the targets from high-priority families were more likely to report ever smoking on the FHTU (85.7% versus 15.6%; \( \chi^2 = 235.3; P < 0.0001 \)), and to have a higher percentage of smokers in the family (74.2% versus 29.6%; \( t (443) = 20.4; P < 0.0001 \)). When compared on measures taken at the initial SMOFAM assessment, targets from high priority families were more likely to report having smoked more than once (43.1% versus 13.4%; \( \chi^2 = 51.2; P < 0.0001 \)), and reported first trying cigarettes at an earlier age (12.9 versus 16.0 years; \( t (468) = -7.0; P < 0.0001 \)). There were no differences between targets from high-priority families and low-priority families on gender or age at the time the FHTU was completed.

**Description of Targets and Family Members**

Table 1 describes key characteristics of the 378 target participants in the IRP and their parents at the initial assessment of the SMOFAM study. The targets, on average, scored in the low to moderately elevated range on the individual environmental and psychosocial measures collected at entry into the study. The generally low levels on the measures of depressive symptoms, stressful events, and conflict with parents are consistent with the selection bias of this sample in which individuals who remained with SMOFAM over the 15 years of the study and who volunteered for participation in the present study had overall lower levels of tobacco dependence behaviors and other risk factors.

Table 2 provides initial data on 867 family members who completed the SHQ. Basic demographics are shown at the top of the table, followed by descriptive statistics grouped according to a phenotypic nomenclature for tobacco use in which simple single-item measures are identified as “Class I phenotypes,” assessments of dependence as “Class II phenotypes,” longitudinal descriptions of the acquisition of tobacco dependence as “Class III phenotypes,” and biological measures of nicotine metabolism as “Class IV phenotypes” (88).

**Demographics and Socioeconomic Status.** Targets and their siblings were 28–29 years of age, whereas their parents averaged between 53 and 55 years of age. Approximately 60% of the targets were female, and 55% of the siblings were male. The prevalence of educational level at a college degree or more was highest among fathers (39.4%) and lowest among siblings (19.9%). A similar pattern was observed for income with fathers being most likely to report making $50,000 or more per year (47.6%) and siblings being least likely to earn that much (20.4%). Overall, whites comprised the largest percentage of participants (85–90%).

**Smoking Behaviors.** Consistent with the initial construction of the SMOFAM cohort and subsequent recruitment of families with a high prevalence of smoking, inspection of Table 2 reveals an overall high prevalence of ever-smoking, ranging from 47.9% among targets to 77.2% among fathers. The prevalence of current smoking also differed widely as a function of generational cohort, with a higher prevalence among targets (29.2%) and siblings (43.9%) than among their mothers (24.7%) and fathers (26.0%). Mothers reported having tried...
their first cigarette at a somewhat later age, 15.3 years, than did fathers, 13.3 years, targets, 13.8 years, and siblings, 13.4 years. The age at which individual family members started to smoke daily was fairly consistent across family members and ranged, on average, from 16.9 to 19.0 years.

**Nicotine Dependence.** In general, targets ($M = 15.8$) and siblings ($M = 16.1$) smoked fewer cigarettes per day when smoking the heaviest than did their mothers ($M = 21.9$) and fathers ($M = 28.5$). Table 2 also reveals higher FTND scores among targets ($4.3$) and fathers ($4.6$) than for targets ($28.5$). Table 2 also reveals higher FTND scores among targets ($28.5$) than for targets ($4.3$) and fathers ($4.6$) than for targets ($28.5$). Table 2 also reveals higher FTND scores among targets ($28.5$) than for targets ($4.3$) and fathers ($4.6$). No clear differences across family members were seen for the Pharmacological or Nonpharmacological subscales of the SMQ (63), or for the extent to which early smoking experiences were pleasant or unpleasant. The reporting of nausea and dizziness in reaction to early exposure to tobacco appeared to be more frequent among targets and siblings than among their parents. The extent to which craving for a cigarette to provide relief from withdrawal was reported often or always appeared to be greater among mothers (38.5%) and fathers (42.5%) than among targets (22.4%) and siblings (26.5%). Across all of the family members, there was a uniformly high percentage of individuals who had ever tried to quit smoking.

**Exposure to Second-Hand Smoke.** The highest estimated number of hours of exposure to second-hand smoke both inside

Table 2. Description of the IRP study sample

<table>
<thead>
<tr>
<th>Item</th>
<th>Target (n = 378)</th>
<th>Mother (n = 166)</th>
<th>Father (n = 127)</th>
<th>Siblings (n = 196)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age when completed questionnaire ($M \pm SD$) &amp; 28.7 ± 1.6 &amp; 53.0 ± 4.6 &amp; 55.4 ± 4.8 &amp; 28.1 ± 5.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (% female)</td>
<td>59.8 &amp; 100.0 &amp; 0.0 &amp; 44.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marital status (% married)</td>
<td>44.2 &amp; 52.4 &amp; 68.5 &amp; 35.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Education (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some college or less</td>
<td>64.8 &amp; 66.3 &amp; 60.6 &amp; 80.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>College degree or more</td>
<td>35.2 &amp; 33.7 &amp; 39.4 &amp; 19.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Annual income (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than $50K</td>
<td>71.0 &amp; 66.1 &amp; 52.4 &amp; 79.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$50K or more</td>
<td>29.0 &amp; 33.9 &amp; 47.6 &amp; 20.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race or ethnicity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>87.6 &amp; 90.4 &amp; 86.6 &amp; 85.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonwhite/mixed/other</td>
<td>12.4 &amp; 9.6 &amp; 13.4 &amp; 14.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Class I phenotypes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ever smoked (% yes)</td>
<td>47.9 &amp; 65.7 &amp; 77.2 &amp; 69.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>52.2 &amp; 34.3 &amp; 22.8 &amp; 30.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimenter (%)</td>
<td>1.9 &amp; 0.6 &amp; 0.0 &amp; 2.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Former (%)</td>
<td>16.7 &amp; 40.4 &amp; 51.2 &amp; 23.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current (%)</td>
<td>29.2 &amp; 24.7 &amp; 26.0 &amp; 43.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age first tried cigarettes ($M \pm SD$)</td>
<td>13.8 ± 4.1 &amp; 15.3 ± 4.2 &amp; 13.3 ± 3.9 &amp; 13.4 ± 3.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age started smoking daily ($M \pm SD$)</td>
<td>16.9 ± 3.8 &amp; 19.0 ± 5.2 &amp; 18.4 ± 6.3 &amp; 17.7 ± 3.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cigarettes smoked per day when smoking heaviest ($M \pm SD$)</td>
<td>15.8 ± 9.1 &amp; 21.9 ± 12.6 &amp; 28.5 ± 17.2 &amp; 16.1 ± 11.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Class II phenotypes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fagerstro Test for Nicotine Dependence ($M \pm SD$)</td>
<td>3.2 ± 2.4 &amp; 4.3 ± 2.8 &amp; 5.1 ± 2.8 &amp; 3.6 ± 2.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking Motives Questionnaire-Pharmacological ($M \pm SD$)</td>
<td>0.7 ± 0.5 &amp; 1.0 ± 0.7 &amp; 1.1 ± 0.7 &amp; 0.8 ± 0.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking Motives Questionnaire-Nonpharmacological ($M \pm SD$)</td>
<td>0.7 ± 0.4 &amp; 0.9 ± 0.4 &amp; 1.0 ± 0.5 &amp; 0.8 ± 0.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early Experiences, nausea (% yes)</td>
<td>18.6 &amp; 13.1 &amp; 12.6 &amp; 21.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early Experiences, dizziness (% yes)</td>
<td>35.3 &amp; 26.3 &amp; 22.3 &amp; 37.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ever tried to quit smoking (% yes)</td>
<td>84.4 &amp; 94.5 &amp; 95.9 &amp; 83.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crave cigarette to provide relief from withdrawal (% yes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>30.9 &amp; 20.2 &amp; 17.0 &amp; 25.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sometimes/often/always</td>
<td>69.1 &amp; 79.8 &amp; 83.0 &amp; 74.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Class III phenotypes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years between first smoking to daily smoking ($M \pm SD$)</td>
<td>3.7 ± 3.6 &amp; 3.4 ± 5.3 &amp; 5.3 ± 6.5 &amp; 4.4 ± 3.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years between first smoking to smoking the most ($M \pm SD$)</td>
<td>7.2 ± 4.7 &amp; 13.8 ± 10.6 &amp; 15.2 ± 10.5 &amp; 7.8 ± 5.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environmental phenotypes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hours of exposure to SHS outside the home ($M \pm SD$)</td>
<td>1.3 ± 2.7 &amp; 0.9 ± 2.3 &amp; 1.5 ± 2.9 &amp; 2.3 ± 3.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hours of exposure to SHS inside the home ($M \pm SD$)</td>
<td>0.5 ± 2.5 &amp; 1.0 ± 3.3 &amp; 1.2 ± 3.5 &amp; 1.1 ± 3.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ancillary phenotypic variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How often drink alcohol (% yes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3–6 days per week to every day</td>
<td>19.6 &amp; 27.4 &amp; 54.4 &amp; 36.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 1 day per week to 1–2 days per week</td>
<td>80.4 &amp; 72.6 &amp; 45.6 &amp; 64.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ever smoked cigars (% yes)</td>
<td>49.6 &amp; 32.5 &amp; 82.7 &amp; 70.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ever used smokeless tobacco (% yes)</td>
<td>35.4 &amp; 3.6 &amp; 36.2 &amp; 43.4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| p | For comparisons across family members $P < 0.0001$. |
| p | For comparisons across family members $P < 0.001$. |
| p | For comparisons across family members $P = ns$. |
| p | For comparisons across family members $P = 0.01$. |
| p | For comparisons across family members $P < 0.05$. |

SHS, second-hand smoke.
and outside the home was reported by the siblings, 3.4 h, followed by fathers, 2.7 h. The number of hours of second-hand smoke exposure reported by the targets and mothers was approximately equal, 1.9 and 1.8 h, respectively. In the first trimester of pregnancy with the target, 36.7% of mothers reported smoking, 30.7% during the second trimester, and 36.7% during the third trimester.

Developmental Tobacco Use Milestones. Fathers reported a greater number of years ($M = 5.3$) between first smoking and daily smoking, whereas mothers reported a smaller number of years ($M = 3.4$). The developmental time course to maximum smoking was greatest for fathers ($M = 15.2$ years) and for mothers ($M = 13.8$ years) than for either targets ($M = 7.2$ years) and siblings ($M = 7.8$ years), most likely because the younger family members are at an earlier point in the developmental trajectory leading to maximal adult dependence.

Ancillary Phenotypes. Fathers in the sample had a higher prevalence of almost daily to daily alcohol consumption (54.4%) than did other family members. The reported prevalence of ever using smokeless tobacco was highest among targets (35.4%), fathers (36.2%), and siblings (43.4%), and relatively low among mothers (3.6%).

Association of Phenotypes with Nicotine Dependence. The bivariate associations between the tobacco-related phenotypes listed in Table 2 and the FTND score were examined for significance in both the targets and siblings, and in the parents. Overall, the pattern of significance was similar for both generations, and subsequent associations are reported for the entire sample. Most of the phenotypes were associated significantly with FTND, thereby supporting the construct validity of many of the various component phenotypes. Phenotypes that were not significantly associated with nicotine dependence included: annual income, $r = -0.06$, age first tried cigarettes, $r = -0.02$, nausea on first exposure to tobacco, $r = -0.03$, frequency of alcohol use, $r = -0.06$, use of alcohol, $r = -0.06$, and of smokeless tobacco, $r = 0.06$. Multiple stepwise linear regression revealed the best set of independent correlates of the FTND in the entire sample to include: education, smoking status (never, experimenter, current, or former), cigarettes per day, reported craving for a cigarette to provide relief from withdrawal, the report of dizziness as an early experience following initial exposure to cigarettes, and the Pharmacological and Nonpharmacological subscales from the SMQ (overall model $r^2 = 0.70$; $F = 76.8; P < 0.0001$).

Discussion

The present article describes an ongoing investigation to address simultaneously the psychosocial and biological risk factors for the development and maintenance of tobacco use. Because of the large public health imperative to seek timely biopsychosocial answers to the prevention of tobacco dependence (28), a strategy of teaming with an existing longitudinal investigation have accelerated the need to adopt a heightened awareness of the numerous ethical, legal, and social issues surrounding the biopsychosocial investigation of complex genetic traits (92, 93). In the case of tobacco dependence, examples already exist of media misrepresentation of findings from both biometric and genetic studies of smoking (94). Changing standards for proper informed consent of both targets and their family members, their right to have access to collected data and to request the destruction of that data, and different standards across institutional review boards and states present unique challenges to an investigation of this nature.

Guiding Principles. Several working principles guided the development of this project. One principle is that tobacco dependence is influenced by an interaction between a complex genetic trait (twin studies reveal that ~50% of the variance in tobacco dependence is attributable to genetic influences; Ref. 89) and environmental influences. Given the multiplicity of determinants to the etiology of tobacco dependence, the availability and utilization of the previous measurements collected by the SMOFAM study is of paramount importance.

Another principle guiding the development of this project is that we do not rely on a single item or set of items to characterize tobacco dependence. In fact, identification of the gold standard for measurement of tobacco dependence is the object of much effort, and previous heavy reliance on unidimensional measurement strategies may well have resulted in a lack of breakthrough findings. The multimodal assessment of behavior has long been recognized as methodologically superior because of the resulting opportunity to evaluate the observed associations against expectations about what should be (convergent validity) and should not be (discriminant validity) associated with the primary constructs of interest (90, 91). Indeed, to our knowledge, the present study is one of only a few studies that have the capability to examine tobacco-related phenotypes ranging from single simple-item measures to biological measures of nicotine metabolism in relation to genotype. The presence of multilevel data will facilitate the testing of several important validity questions. For example, the 3HC: cotinine ratio, as a marker of CYP2A6 activity, should demonstrate a robust association with allelic variation in CYP2A6 and with fewer adverse reactions to initial consumption of tobacco (convergent validity) and should have little or no association with allelic variation in DRD2 and with measures of depressive symptomatology (discriminant validation).

A third principle underlying the present investigation is that there is need for a close connection between the observed epidemiology of tobacco use, its aggregation within families, and the application of molecular genetic strategies to establish phenotype-genotype associations in those families. The investigation of tobacco dependence within families began in this study with the collection of the family history of tobacco use and then continued with the use of this information to identify families that have a high likelihood of being informative in the pursuit of phenotype-genotype associations. Biometric tools are being used to identify patterns of familial aggregation of tobacco dependence within these families, and these will then be used to guide molecular genetic analyses in the identification of specific markers that may be important in the development of tobacco dependence.

A fourth principle underlying the work is that any investigation of genetic risk factors for tobacco dependence must be cognizant of the social/political context in which the work is being performed. Developments from the human genome project have accelerated the need to adopt a heightened awareness of the numerous ethical, legal, and social issues surrounding the biopsychosocial investigation of complex genetic traits (92, 93). In the case of tobacco dependence, examples already exist of media misrepresentation of findings from both biometric and measured genetic studies of smoking (94). Changing standards for proper informed consent of both targets and their family members, their right to have access to collected data and to request the destruction of that data, and different standards across institutional review boards and states present unique challenges to an investigation of this nature.

Generalizability of the Sample. Relative to the original SMOFAM sample, the volunteers for the family history portion of this genetic investigation in families are more likely to be female, younger, and less likely to have smoked more than once. Relative to those who completed a FHTU questionnaire and who did not continue participation, those who also com-
pleted the SHQ were similar with respect to smoking and quitting history but more likely to be female and historically involved in the repeat measurements of SMOFAM. Overall, then, it would appear that the sample included in this genetic investigation is biased away from individuals with severe dependence and the prevalence of risk factors such as family conflict. Most likely, then, the present investigation involves a sample that is best described as community-based and less severely dependent in nature.

**Planned Genetic Analyses.** To evaluate the role of Mendelian inheritance in the familial clustering of tobacco-related phenotypes, complex segregation analysis (95, 96) will be performed using likelihood procedures that test for a transmission model that best fits the data. Two types of measured genetic analyses are planned for this study: linkage segregation analysis (97–99) and family based allelic association analysis for candidate genes (100–103). The preferred chromosome segregation analysis method, referred to as the variance component method, allows for the inclusion and testing for the significance of environmental covariates. PEDCHECK (104) and SIMWALK2 (78) will be used to identify probable genotype errors using both SNP and microsatellite genotypes. Family-based haplotype association analysis (105, 106) will be used to identify alleles that are associated with tobacco-related phenotypes. Gene-tobacco use phenotype associations, gene × gene, and gene × environment interactions will be examined using newly developed methods (107–109).

**Analysis of Genetic and Environmental Risk Factors.** Latent growth curve modeling (110, 111) and latent growth mixture modeling (112–115) will be used to examine the combined effects of genotype status and other risk factors as predictors of the rate of progression toward regular smoking (i.e., as a function of slope, defined by multiple measures over time) and tree-structured survival analysis (116–118) to examine relationships between allele status for candidate susceptibility genes (coded as present or absent) and environmental risk factors (coded as present or absent), with age at onset of regular smoking serving as the time-based outcome variable.

The interdisciplinarity approach is fundamentally a systems-oriented view of tobacco dependence in which members of multiple disciplines seek to develop an understanding about how changes in one component of the “super-system” of tobacco dependence alters the status of another subsystem (119). Whereas the application of this level of integration seems well placed in the field of tobacco dependence (120), its implementation is novel, and it is still too early to determine whether it will lead to breakthrough findings. However, it should be noted that a number of examples of successful interdisciplinary efforts exist within the field of medicine (e.g., cardiovascular health and behavior, and schizophrenia; Ref. 121); imaging in clinical medicine, the development of synthetic biomaterial and artificial organs (122), thereby demonstrating the real-world utility of integrative, systems-oriented research. Pellmar and Eisenberg (121) identify several prime areas for future integration and highlight the scientific foundation. Nicotine Tob. Res., 4: S11–S33, 1999.


Environmental and Genetic Determinants of Tobacco Use: Methodology for a Multidisciplinary, Longitudinal Family-Based Investigation


Updated version
Access the most recent version of this article at:
http://cebp.aacrjournals.org/content/12/10/994

Cited articles
This article cites 81 articles, 6 of which you can access for free at:
http://cebp.aacrjournals.org/content/12/10/994.full#ref-list-1

Citing articles
This article has been cited by 3 HighWire-hosted articles. Access the articles at:
http://cebp.aacrjournals.org/content/12/10/994.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.