

Genome-wide Linkage Scan to Identify Loci for Age at First Cigarette in Dutch Sibling Pairs

Jacqueline M. Vink,^{1,4} Danielle Posthuma,¹ Michael C. Neale,² P. Eline Slagboom,³
and Dorret I. Boomsma¹

Received 1 Apr. 2005—Final 25 Jul. 2005

The heritability of age at first cigarette was estimated in 5883 Dutch twins and siblings registered with the Netherlands Twin Register. Heritability was 60% for males and 39% for females. Shared environmental influences were found in females only (30%). Linkage analyses were performed on data of 422 DZ twins and siblings from 175 families, forming 368 sibling pairs. Genomic regions that may harbor susceptibility loci for age at first cigarette with LOD score greater than 2 were detected on chromosomes 5, 14 and 22. A simultaneous analysis of these three genomic regions showed that most of the variance was explained by the linkage effect on chromosome 5 (205 cM). This peak encloses the D1A dopamine receptor gene which is a functional candidate gene for smoking behavior.

KEY WORDS: Age at first cigarette; dopamine; linkage; QTL; sex differences.

INTRODUCTION

Most smokers begin smoking during adolescence and early age of smoking initiation is related to more frequent current smoking, daily smoking and more dependent smoking (Everret *et al.*, 1999; Lando *et al.*, 1999). Smoking initiation (yes/no) is influenced by both genetic and environmental influences (Li *et al.*, 2003a; Sullivan and Kendler, 1999). However, relatively little is known of the genetics of age at onset of smoking. A study by Heath *et al.* (1999) reported a heritability of 62% in a young cohort (age ≤ 30 years) and 51% in an older cohort (age > 30 years). The

remaining variance was attributed to unique environmental influences. There was no evidence for shared environmental influences on age at onset of smoking. In contrast, Stallings *et al.* (1999) reported that recollected age at first use and age at daily use was not influenced by genetic factors but only by shared environmental and unique environmental influences. The age of the participants in their study ranged from 50 to 96 years (Stallings *et al.*, 1999). Both studies used a different analytical approach, a different definition of the variable and different age cohorts, which may account for the different results.

Several genome wide scans of smoking behavior have been published (Bergen *et al.*, 1999; Bierut *et al.*, 2004; Duggirala *et al.*, 1999; Gelernter *et al.*, 2004; Li *et al.*, 2003b; Saccone *et al.*, 2003; Straub *et al.*, 1999; Sullivan *et al.*, 2004; Vink *et al.*, 2004; Wang *et al.*, 2005) but none of these scans used the phenotype “age at first cigarette”. A recent association study in a Chinese population used the related phenotype “early smoking onset” and has indicated a possible association between this phenotype and polymorphisms of the dopamine transporter gene on chromosome 5p (Ling *et al.*, 2004). The rs2072-A

¹ Department of Biological Psychology, Vrije Universiteit, BT Amsterdam, The Netherlands.

² Departments of Psychiatry and Human Genetics, Virginia Institute for Psychiatry and Behavioral Genetics, Medical College of Virginia, Virginia Commonwealth University, Richmond, VA, USA.

³ Section Molecular Epidemiology, Sylvius Laboratory, Leiden University Medical Centre, Leiden, The Netherlands.

⁴ To whom correspondence should be addressed at Department of Biological Psychology, Vrije Universiteit, Van der Boerhorststraat 1, 1081, BT Amsterdam, The Netherlands. Tel.: +31-20-4448787; Fax: +31-20-4448832; e-mail: jm.vink@psy.vu.nl

allele was significantly associated with smoking onset less than 18 years (OR = 4.0 in the total population of 668 smokers).

In summary, less is known about the genetic influences on age at first cigarette and results are contradictory. In the present paper we will first investigate whether age at smoking the first cigarette is a heritable trait in a Dutch twin-family sample. We also tested whether age at first cigarette is a unidimensional trait. Next we report the results from a complete genome scan of age at first cigarette in a genotyped subsample of dizygotic twins and sibling pairs.

METHODS

Subjects

This study is part of an ongoing twin-family study on health, lifestyle and personality of the Netherlands Twin Register (Boomsma *et al.*, 2002). For this study data from the surveys of 1993, 1995 and 1997 were used. In total, 7836 twins and siblings participated at least in one of those surveys. Participants aged 18 years and older when completing the survey were selected ($n=6175$). No data on age at first cigarette were available for 74 subjects. Half-siblings and adoption-siblings were excluded ($n=47$), leaving 6054 participants in the study. The mean age when completing the survey was 26.2 (SD 9.2) for males and 26.3 (SD 9.3) for females.

For the genetic model fitting analyses the twins and at most two brothers and two sisters were selected. This procedure excluded 18 brothers and 39 sisters. The remaining sample consisted of 3687 twins (587 MZM, 470 DZM, 1041 MZF, 670 DZF, 453 DOS females and 466 DOS males), 758 brothers and 894 sisters from 2453 families. DZ twin and non-twin sibling pairs were selected for the linkage scan. DNA marker data were available for 630 sibling pairs. Genotypic and phenotypic data were available for 422 DZ twins and siblings from 175 families, forming 368 sibling pairs. The sample consisted of 117 sister pairs, 95 brother pairs and 156 opposite sex pairs.

Phenotype

Twins and siblings were asked at what age they smoked their first cigarette. Response categories were: 11 years or younger, 12, 13, 14, 15, 16, 17, 18 years or older, and never. This question was included in the surveys of 1993, 1995 and 1997. Our

analyses rely on the retrospective recall of age at smoking first cigarette. Because self-reported data on substance use can be unreliable (Johnson and Mott, 2001), we selected subjects on the following criteria: (1) not more than 1 year difference for age at first cigarette in the longitudinal data or (2) subjects who reported they never smoked in the first survey(s) and who reported an age at first cigarette in later survey(s). According to those criteria, 464 subjects (about 16% of all subjects with longitudinal data) were excluded.

Genotyping

Genotyping was performed by two facilities: the Mammalian genotyping service, Center for Medical Genetics, Marshfield, USA, and the Molecular Epidemiology Section, Leiden University Medical Centre, The Netherlands. Ninety four persons were included in both scans. A 379 autosomal marker genome scan (9.44 cM spacing) was done by the Mammalian Genotyping Service, using microsatellite screening set 10 (Yuan *et al.*, 1997) with few alternative markers. Markers with excessive recombination rates ($n=5$) and inconsistently marked markers ($n=5$) were not included in the analysis. A 419 marker genome scan (8.34 cM spacing) was performed by the Molecular Epidemiology Section, Leiden University Medical Centre, The Netherlands (Beekman *et al.*, 2003; Heijmans *et al.*, 2005).

Mendelian errors were detected using PEDSTATS and unlikely double recombinants using MERLIN; both types of error were removed using PEDWIPE (Abecasis *et al.*, 2002). Pedigree relationships in the entire sample were checked with the GRR program (Abecasis *et al.*, 2001). The location of the markers was taken from an integrated genetic map with interpolated genetic map positions (<http://www2.qimr.edu.au/davidD/>). The position is in deCODE cM (Kong *et al.*, 2002), estimated via locally weighted linear regression from the Build 34.3 (and 35.1) physical map positions and published deCODE and Marshfield genetic map positions. If available, parents were typed for between 344 and 375 markers (mean of 363 ± 6). For offspring, the number of typed markers ranged from 344 to 678, with an average of $389 (\pm 69)$ total markers.

Genetic Model Fitting

Structural equation modeling in Mx was used to fit models to the data (Neale *et al.*, 1999). First, to

explore whether the variable age at onset of the first cigarette (measured with the categories 11 years or younger, 12, 13, 14, 15, 16, 17, 18 years or older, never) is unidimensional we fitted three different models to the data (Heath and Martin, 1993; Koopmans *et al.*, 1999; Vink *et al.*, 2005):

- *Single liability model*: postulates that the liability to age at onset of the first cigarette is unidimensional.
- *Independent liability model*: assumes two independent liability dimensions for smoking initiation and the age at first cigarette.
- *Combined model*: the combined model includes features of both the single liability model and the independent liability model. Like the independent model it postulates the existence of two separate dimensions but it allows the possibility that there are some genetic and environmental risk-factors which influence both dimensions.

In a saturated model we tested whether the means and variances of males and females differed from each other, and whether the covariances of twin–sib and sib–sib pairs were different. We also tested whether age at completing the survey influenced the individual differences in reported age of smoking the first cigarette by including “age at completing the survey” as a covariate in the model fitting procedure. The resulting most parsimonious model was the one which was used to test the relative contributions of genetic and environmental influences. Sources of variation that were considered in modeling were additive genetic variance (σ_{am}^2 for males and σ_{af}^2 for females) shared environmental variance (σ_{cm}^2 for males and σ_{cf}^2 for females), and variance due to non-shared environmental influences including error of measurement (σ_{em}^2 for males and σ_{ef}^2 for females) (Fulker and Cherny, 1996).

Under this model, the expected covariance between sibling pairs is given by

$$\begin{aligned} & r_g \sigma_{am}^2 + r_c \sigma_{cm}^2 \text{ for brother pairs,} \\ & r_g \sigma_{af}^2 + r_c \sigma_{cf}^2 \text{ for sister pairs and,} \\ & r_g \sigma_{am} \sigma_{af} + r_c \sigma_{cm} \sigma_{cf} \text{ for brother–sister pairs.} \end{aligned}$$

where r_g is 1 for MZ twin pairs, 0.5 for same-sex sibling pairs and is estimated between 0 and 0.5 for opposite sex sibling pairs (to test for qualitative sex differences), r_c is 1 in MZ and DZ twin pairs and estimated between 0 and 1 in the non-twin sibling pairs (twin–sib or sib–sib), to test for a special shared twin environment.

The means were modeled according to the formula:

$$Y_{ij} = \mu + \beta \text{Age}_{ij} + e_{ij}$$

where y_{ij} is the observed phenotype for sibling j in the i th family, μ denotes the grand mean and was allowed to differ for males and females, β represents the regression coefficient for age, age_{ij} represents the age of sib j from the i th family, and e_{ij} represents the residual term that is not explained by the fixed effects of age and sex.

IBD Estimates

The probability of sharing 0, 1 or 2 alleles IBD were estimated for a 1 cM grid using the Lander–Green algorithm implemented in the program Merlin (Abecasis *et al.*, 2002). The estimate of the proportion of alleles shared identical by descent ($\hat{\pi}$) was obtained as: $\hat{\pi}_{ijk} = 0.5p_{(IBD=1)ijk} + 1p_{(IBD=2)ijk}$ where $\hat{\pi}_{ijk}$ is the estimated proportion of alleles shared IBD between sib j and k for the i th family, and $p_{(IBD=1)ijk}$ and $p_{(IBD=2)ijk}$ are the probabilities that sib j and k share 1 or 2 alleles, respectively, conditional on the marker information (Sham, 1998).

Linkage Analyses

Linkage analyses were performed in Mx (Neale *et al.*, 1999). Variance components linkage analyses included the variation explained by a putative quantitative trait locus (Q) at or near a specific location on the genome. Estimates of this variance component were obtained using the $\hat{\pi}$ approach, in which the covariance due to the marker or trait locus for a sib pair is modeled as a function of the IBD status of the sib pair. The variance in liability and the phenotypic variance of the residual term, respectively, were decomposed into additive genetic variance (σ_a^2), shared environmental variance (σ_c^2), variance due to non-shared environmental influences (σ_e^2) and variance due to the QTL (σ_q^2) (Fulker and Cherny, 1996).

$$\begin{aligned} & r_g \sigma_{am}^2 + r_c \sigma_{cm}^2 + \hat{\pi}_{ijk} \sigma_{qm}^2 \text{ for brother pairs,} \\ & r_g \sigma_{af}^2 + r_c \sigma_{cf}^2 + \hat{\pi}_{ijk} \sigma_{qf}^2 \text{ for sister pairs and,} \\ & r_g \sigma_{am} \sigma_{af} + r_c \sigma_{cm} \sigma_{cf} + \hat{\pi}_{ijk} \sigma_{qm} \sigma_{qf} \text{ for brother–sister pairs} \end{aligned}$$

where r_g and r_c and $\hat{\pi}_{ijk}$ are defined as above. The analyses included also phenotypic data from the (non-genotyped) MZ twin pairs to allow the distinction between background additive genetic and other familial effects. For the MZ twins the covariances were modeled as

$$\sigma_{am}^2 + \sigma_{cm}^2 + \sigma_{qm}^2 \text{ for MZM and,}$$

$$\sigma_{af}^2 + \sigma_{cf}^2 + \sigma_{qf}^2 \text{ for MZF}$$

First, male and female QTL effects were estimated as separate parameters. Minus twice the log-likelihood of this model compared to a model where the male and female QTL components are constrained equal gives a test for sex-specific QTL effects. When both the male and female QTL effects are simultaneously dropped from the model, the difference in degrees of freedom is 2. The distribution of the test-statistic is in this instance: $1/4\chi^2_0 : 1/2\chi^2_1 : 1/4\chi^2_2$. One alternative is to treat it as a standard two degrees of freedom test, which is more conservative. We have decided to use this latter, more conservative method. Finally, the QTL effect for males was equated to the QTL effect for females and linkage was tested by comparing the fit of this model with a model without a QTL (1 df). Significance of genetic variation due to the QTL was evaluated by the likelihood ratio test, from which the LOD score can be calculated by dividing the test statistic χ^2 by $2 \ln 10$ (~ 4.6) (Sham, 1998).

The three loci with LOD scores > 2.0 were modeled simultaneously in a three-locus linkage analysis. This simultaneous analysis is more robust to the well-known overestimation of variance attributable to markers in a linkage analyses. In a single-locus analysis, the QTL effects may be overestimated and together sum to a percentage of explained variance $> 100\%$ (Geus *et al.*, 2005; Goring *et al.*, 2001). The three-locus analysis constrains the sum of the QTL variances to be less than 100% of the phenotypic variance. To calculate a LOD-score for the model without the three loci, the p -value of the χ^2 difference with 3 degrees of freedom was used to estimate the corresponding χ^2 with 1 degree of freedom. This χ^2 was used to calculate the LODscore by dividing this test statistic by $2 \ln 10$.

For the chromosome with the most promising finding, the chromosome-wide empirical p -value was determined by simulation, using the 1000 \times permutation algorithm as described by Lystig (2003) and implemented (for genome-wide thresholds) by Posthuma *et al.* (2005). Permuted datasets were obtained as follows: each row in the observed dataset represents one family and contains phenotypic data for each individual within that family as well as the IBD probabilities across the chromosome for all pairs within that family. This file is split into a phenotypic file and an IBD file. Families in the experiment are labeled with unique numbers one through n . The phenotypic data are then shuffled by taking a random

permutation of the indices $1, \dots, n$ and matching the i th phenotypic trait value to the family with index given by the i th element of permuted indices. This permuted vector of traits is matched with the original (non-permuted) genotypic information for all families. One thousand permuted datasets were generated. The significant linkage threshold was defined as the LOD-score occurring in 50 of the 1000 permutation corresponding to a probability of 0.05 in a genome scan (Lander and Kruglyak, 1995).

RESULTS

Descriptives

The distribution of age at first cigarette in males and females is shown in Figure 1. The distribution is skewed due to the number of subjects that never smoked a cigarette (around 30% of the sample). The skewness statistic is -0.402 with SE 0.033. In general, slightly more males than females try their first cigarette on an early age. The median age is 15 years for males and 16 years for females.

To test whether the variable was unidimensional three different models were fitted to the data: a single liability model, an independent liability model and a combined model (Heath and Martin, 1993; Vink *et al.*, 2005). The single liability model, which postulates that the liability to age at onset of the first cigarette is unidimensional, was the most parsimonious model (Table I). Thus, subjects who never tried a cigarette could be included in the analyses.

Twin and Sibling Correlation

The means for males and females were significantly different; males report an earlier age at onset than females (16.3 for males vs. 16.7 for females). The variances of age at first cigarette could be constrained to be equal for males and females. The covariances between same sex twin-sibling pairs and same-sex sibling-sibling pairs and the covariances between opposite-sex twin-sibling pairs and opposite-sex sibling-sibling pairs could also be constrained to be equal. Dropping the age-effect (age at completing the survey) significantly worsened the fit of the model. Table II shows the correlations and 95% CI for twin and sibling pairs for the age at smoking the first cigarette. For males the MZ correlations are twice as high as the DZ correlations, suggesting additive genetic influences. For females, the DZ correlation was greater than half the MZ correlation, indicating that

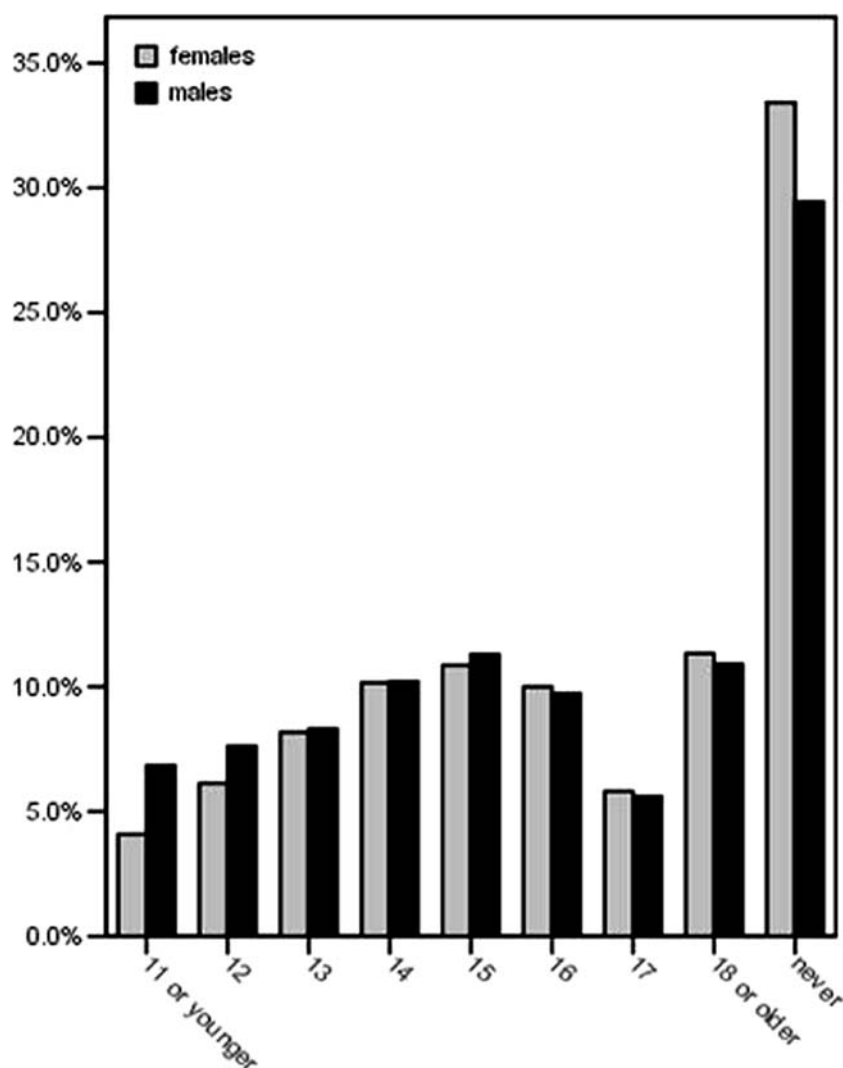


Fig. 1. Distribution of the age at first cigarette for males and females.

common environmental effects may also contribute to the familial clustering of individual differences in age at smoking the first cigarette. The DZ twin correlation was greater than the sibling correlation suggesting a special twin environment.

Genetic Model Fitting

Model fitting results for the age at smoking the first cigarette are summarized in Table III. The models included sex-differences in the mean and in parameter estimates. Setting the shared environmental correlation in non-twin sibling pairs to 1 (model 2) significantly reduced the goodness of fit of the model indicating a special twin environment. Model 3 tested

whether the genetic correlation (r_g) in opposite sex siblings could be constrained to be 0.5, like the correlation in same sex sibling pairs. Constraining this correlation significantly worsened the fit of the model. For males the MZ correlation was (almost) twice as high as the DZ correlation, suggesting that shared environmental influences are absent or very low. Dropping the shared environment from the model (model 4) did not reduce the fit of the model in males, but resulted in a significant ($p < 0.001$) loss of fit in females (model 5). In conclusion, the variation in the age at first cigarette was best described by an AE model for males ($A = 60\%$, 95%CI 51–68% and $E = 40\%$, 95%CI 34–47%) and an ACE model for females ($A = 39\%$, 95%CI 22–59%, $C = 30\%$, 95%CI

Table I. Goodness-of-fit of the Single Liability Dimension Model, the Independent Liability Model and the Combined Model to the Data on Age at First Cigarette

	χ^2	df	AIC
Single liability model	587.78	383	-178.21
Independent liability model	722.03	378	-33.98
Combined model	625.69	376	-126.31

df=degrees of freedom, $AIC = \chi^2 - 2df$, this is a measure of the parsimony of the model, a lower AIC indicates a more parsimonious model.

Table II. Twin- and Sibling-Correlations with 95%CI for Age at First Cigarette

	<i>r</i>	95%CI	<i>n</i> pairs
MZM	0.59	0.50–0.65	245
DZM	0.36	0.24–0.46	186
MZF	0.69	0.66–0.74	456
DZF	0.50	0.41–0.63	274
DOS	0.23	0.08–0.36	364
Male sib pairs	0.23	0.16–0.30	544
Male–female sib pairs	0.18	0.11–0.24	1303
Female sib pairs	0.31	0.23–0.38	812

MZM=monozygotic male twins, DZM=dizygotic male twins, MZF=monozygotic female twins, DZF=dizygotic female twins, DOS=dizygotic opposite sex twins, male sibs pairs=brother–brother pairs (including a male twin with a brother), female sib pairs=sister–sister pairs, male–female sib pairs=brother–sister pairs.

11–46% and $E = 31\%$, 95%CI 27–46%). The genetic correlation in the opposite sex siblings was estimated to be 0.40 (95%CI=0.27–0.58) in the AE–ACE model. The shared environmental correlation in sibling-pairs (non-twin siblings) was estimated to be 0.33 (95%CI=0.16–0.57). The effect of age was rather small ($\beta = -0.017$ in AE–ACE model) but significant ($\chi^2 = 12.7$, $df = 1$, $p = 0.0004$).

Linkage Analyses

Because variation in age at first cigarette was heritable, we explored this phenotype in a linkage analysis. The linkage analyses were performed on a smaller sample. As summarized above, we found that the genetic correlation in opposite sex sibling pairs was somewhat lower than 0.5, therefore the male and female QTL effects were estimated as separate parameters. To test for sex-specific QTL effects, minus twice the log-likelihood of this model was subtracted from that obtained when the male and female QTL effects were equated. For most positions along the genome, constraining the QTL effects for males and females to be equal did not worsen the fit of the model. For only five regions, the model with sex-specific QTL effects provided a more parsimonious model than the model with a single QTL effect. For four of these regions a model without QTL effects did not fit worse, suggesting no significant linkage for these regions. However, for a region on chromosome 6 (166–182 cM), the model with sex-specific QTL effects was more parsimonious, and the LOD score in that region was higher (1.3) than for the model with 1 QTL (0.5). The grey line in Figure 2 (chromosome 6) represents the LOD-scores for a model with sex-specific QTL effects.

Thus, the QTL effect for males was equated to the QTL effect for females and linkage was tested by comparing the fit of this model to a model without a QTL (Fig. 3). LOD scores greater than 2 were found for chromosomes 5, 14 and 22 (Fig. 4). Three other peaks were found with LOD scores between 1.5 and 2.0 on chromosomes 6, 7 and 10. The six regions with LOD score >1.5, the highest LOD score and the location of the highest LOD score are shown in Table IV.

The three loci with LOD scores >2.0 were modeled simultaneously in a three-locus linkage analysis. For all three loci simultaneously, an

Table III. Model Fitting Results for Age of Smoking First Cigarette

		-2LL	df	χ^2	Δdf	<i>p</i>
1	ACEm, ACEf, $r_{csibs} = \text{free}$, $r_{gos} \text{ free}$	24818.081	5321			
2	ACEm, ACEf, $r_{csibs} = 1$, $r_{gos} \text{ free}$	24831.646	5322	13.566	1	0.000
3	ACEm, ACEf, $r_{csibs} = \text{free}$, $r_{gos} = 0.5$	24825.957	5322	7.877	1	0.005
4	AEm, ACEf, $r_{csibs} = \text{free}$, $r_{gos} \text{ free}$	24819.693	5322	1.613	1	0.204
5	AEm, AEf, $r_{csibs} = \text{free}$, $r_{gos} \text{ free}$	24831.748	5323	12.055	1	0.000

Subjects are aged 18 years or older. Age was included as covariates. The means for males and females were estimated separately. m=males, f=females, r_{gos} =genetic correlation in opposite sex sib–sib and twin–sib pairs, r_{csibs} =shared environmental correlation in sib–sib and twin–sib pairs.

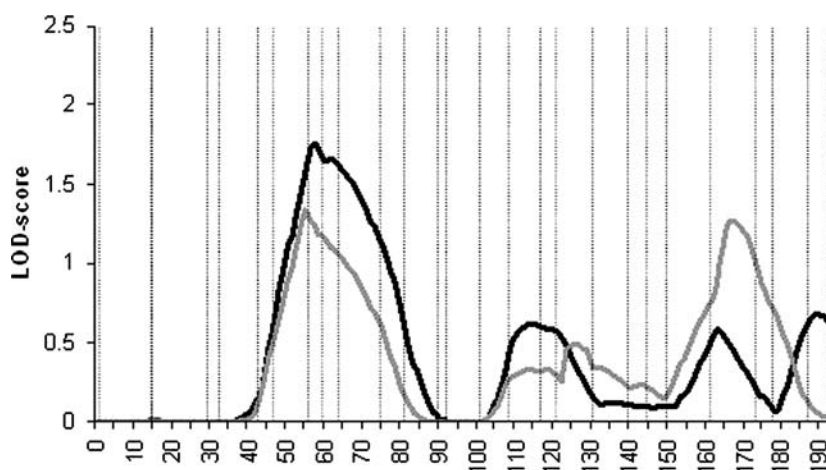


Fig. 2. Linkage results for sex-specific linkage (grey line) and linkage results for analyses where QTL effects were equated for males and females (black line). For the region from 166 to 182 cM the sex-specific model were more parsimonious. LOD-scores of the sex-specific linkage model were higher in this region than LOD-scores of the model without sex-specific QTL effects. For the rest of the chromosome, the model where QTL effects were the same for males and females did not fit significantly worse.

adjusted LOD score of 1.8 was found (Table V). For chromosome 22, the evidence of linkage disappeared in the multilocus analysis. For chromosomes 5 and 14 the evidence weakened in comparison with the single-locus analyses. The peak on chromosome 5 explained the most of the variance in age at first cigarette.

The significant linkage threshold was defined for chromosome 5 by simulation, using a 1000× permutation algorithm. The LOD score occurring in 50 of the 1000 permutations (corresponding to a probability of 0.05) was 2.01. The empirical p -value for a LOD score of 2.09 was 0.04, indicating that the peak on chromosome 5 is significant.

DISCUSSION

A quantitative trait representing the age at first cigarette was assessed in a genetic and linkage analyses. Correlations between sibling pairs were consistent with the existence of a genetic component, as were variance components estimates. The linkage analyses showed suggestive linkage on chromosomes 5, 14 and 22.

Variation in age at first cigarette was significantly influenced by genetic factors, heritability was estimated at 60% for males and 39% for females. Shared environmental influences were found only in females (30%).

Heath *et al.* (1999) explored a slightly different phenotype: onset of smoking, based on the questions “have you ever been a smoker” and if yes “at what

age did you start smoking”. A one-dimensional model gave a poor fit to those data. In the study of Heath *et al.* (1999), the age at onset was largely influenced by genetic factors (62% in young cohort and 51% in older cohort) but not by shared environmental factors. In a study of Stallings *et al.* (1999) the variable “At what age did you smoke your first cigarette” was analyzed. The genetic analyses were performed on residual scores of the age at onset variable, partialing out effects of gender, age and interactions between these effects. Dropping the additive genetic influences from the model did not result in a significant decrement in model fit. The estimate of shared environmental influences was 42%. It should be noted that this sample was an older cohort (born between 1904 and 1937) and that the environmental culture regarding smoking has changed considerably since the time subjects in this sample initiated smoking. In addition, 75% of the sample was female and the genetic analyses did not allow for sex differences.

The genetic analyses in the present study were performed on a relatively young sample (mean age 26 years). The correlation between the genetic factors in opposite-sex siblings was estimated at 0.40, suggesting partly different genes are expressed in males and females. Therefore the linkage model allowed for sex-limitation. Sex-limited effects on the phenotypic variance, also known as genotype-by-sex interaction, has been detected in several physiological and psychological traits (Comuzzie *et al.*, 1993; Lamb *et al.*,

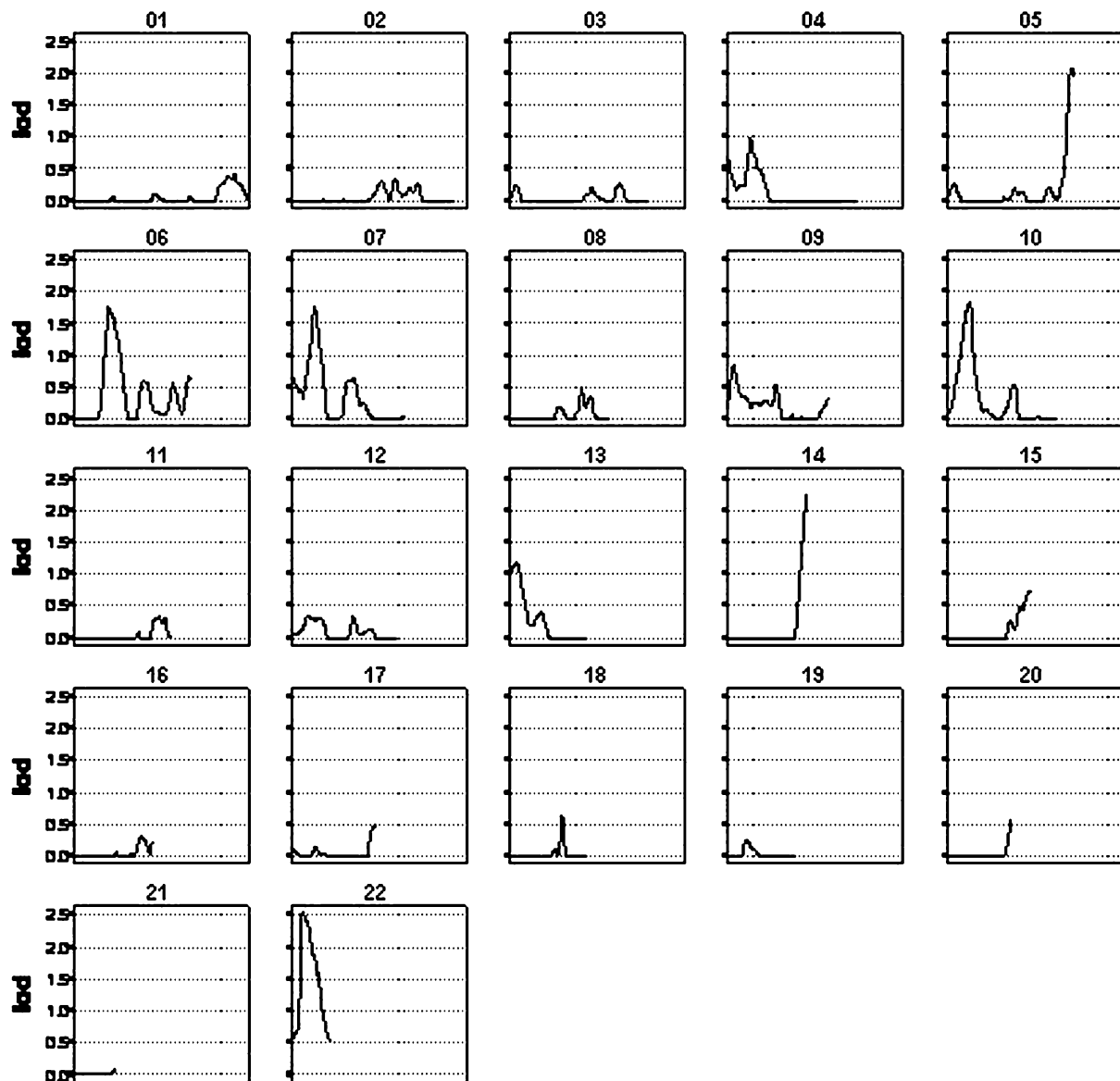


Fig. 3. LOD scores across the genome (chromosomes 1–22) for age at first cigarette. centiMorgans are shown on the x -axis and LOD-scores are shown on the y -axis.

2005; Towne *et al.*, 1993; Weiss *et al.*, 2005). The implementation of models in which the effects of quantitative trait loci are allowed to differ between males and females has been described previously (Towne *et al.*, 1997). Comparison of the model where QTL effects for males and females were allowed to differ and the model where QTL effects for males and females were the same showed that the latter was not fitting worse. An exception was a region on chromosome 6. For the region from 166 to 182 cM the model where the QTLs for males and females were allowed to differ was more parsimonious. This re-

sulted in a higher LOD score (1.3 vs. 0.5) for that specific region.

Results of the model where the effects of the QTL on males and females were the same showed suggestive linkage peaks (with a LOD score higher than 2) on chromosomes 5, 14 and 22.

On chromosome 5, a peak was located between 196 and 209 cM. The simultaneous three-locus analyses confirmed the importance of this peak. Two other studies detected a peak in this region (Duggirala *et al.*, 1999) detected a peak with a LOD score of 3.2 for “ever smoked” on 217 cM while (Saccone *et al.*,

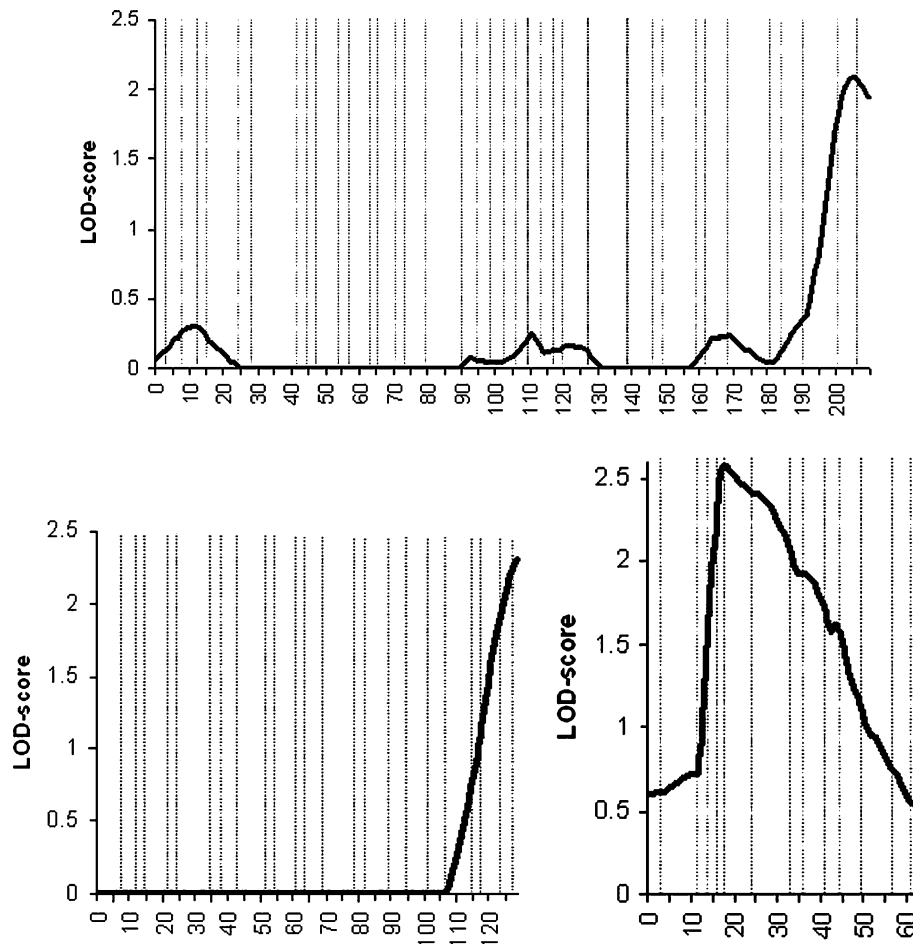


Fig. 4. Linkage results for age at first cigarette on chromosomes 5, 14 and 22. Distance in cM is shown along the x-axis and the LOD-scores along the y-axis.

Table IV. Highest LOD Scores (≥ 1.5) and Locations

Chromosome	Location (LOD ≥ 1.5)	Highest LOD-score	Location highest LOD	Markers	Marker location
5	196–209	2.09	205	ATA52D02 D5S408	200.6 208.2
6	55–67	1.75	57	D6S1051 D6S1610	56.8 60.1
7	34–42	1.77	36	D7S3051 D7S1802	32.6 36.2
10	28–40	1.84	36	D10S1430 D10S1653	32.2 38.6
14	120–128	2.30	128	D14S292 D14S1007	124.9 127.9
22	13–44	2.50	20	D22S345 D22S315	18.8 23.4

2003) detected a peak with LOD score of 1.10 for maximum number of cigarettes (non-smokers excluded) on 200 cM. Two other studies reported link-

age on chromosome 5 in a non-overlapping region. A peak was found for nicotine dependence at 160 cM (Sullivan *et al.*, 2004) and a peak was found for

Table V. Simultaneous Linkage Analysis of Three Loci with a Single-Locus LOD Score >2.0 for Age at First Cigarette

Model	Sign test for:		% contribution to variance						Δdf	χ^2	LOD
			A	C	E	Chr 5	Chr 14	Chr 22			
ACEf, AEm +	–	m	0.10	–	0.42	0.33	0.15	0			
Chr 5, 14, 22	–	f	0.01	0.21	0.30	0.33	0.15	0			
ACEf, AEm +	Chr 5	m	0.44	–	0.42	–	0.15	0	1	6.691	1.45
Chr 14, 22		f	0.21	0.33	0.31	–	0.15	0			
ACEf, AEm +	Chr 14	m	–0.14	–	0.42	0.44	–	0	1	4.154	0.90
Chr 5, 22		f	0.02	0.22	0.30	0.44	–	0			
ACEf, AEm +	Chr 22	m	0.10	–	0.42	0.33	0.15	–	1	0	0
Chr 5, 14		f	0.01	0.21	0.30	0.33	0.15	–			
ACEf, AEm	Chr	m	0.58	–	0.42	–	–	–	3	13.11	1.76
	5, 14, 22	f	0.38	0.31	0.31	–	–	–			

The percentage contribution to the variance is given for background genetic influences (A), shared environment (C), environmental influences (E), and the three loci on chromosomes 5 (205 cM), 14 (128 cM) and 22 (20 cM). The significance of the contribution of each locus separately and of all three loci together derives from the contrast of the fit of the model with 3 loci (model 1) and the fit of the models without one of the QTL effects (models 2–4) or without all QTL effects (model 5). Chr = chromosome(s); Δdf = change in degrees of freedom, f = females, m = males.

habitual smoking at 119 cM (Bierut *et al.*, 2004). The peak found in the present study encloses a gene coding for the D1A dopamine receptor. Dopaminergic genes are likely candidates for heritable influences on cigarette smoking (Lerman *et al.*, 1999; Sabol *et al.*, 1999). The dopamine D1 receptor gene might play a role in addictive behaviors. In a group of controls and individuals attending a smoking cessation clinic, 39.3% of the controls vs. 66.1% of the smokers was homozygous for the 11 or 22 genotype of the dopamine receptor D1 gene ($p=0.0002$), (Comings *et al.*, 1997). This region on chromosome 5 merits follow-up in future studies.

The peak found on chromosome 14 (120–128 cM) is located in the same region as a peak for nicotine dependence on 126 cM (Sullivan *et al.*, 2004). Three other studies found peaks in the neighborhood of this region: on 111 cM for smoking rate (Li *et al.*, 2003b), on 95–110 cM for smoking initiation (Bergen *et al.*, 1999) and on 87 cM for maximum number of cigarettes (Saccone *et al.*, 2003).

The highest LOD-score was found for chromosome 22. Remarkably, in the simultaneous linkage analyses of the loci on chromosomes 5, 14 and 22, the peak on chromosome 22 was not significant anymore. No other genome-scan for smoking behavior detected linkage on this chromosome. The highest LOD-score (2.5) was found at 20 cM but the peak covered most of the chromosome. The CYP2D6 gene (a cytochrome p450 enzyme) is located on chromosome 22, but is not located in the same region as our highest peak. The cytochrome p450 enzymes are involved in

nicotine metabolism and neurotransmitter synthesis. Both CYP2A6 (on chromosome 19) and CYP2D6 (on chromosome 22) catalyze the formation of cotinine from nicotine. Taken together, the peak on chromosome 22 might be spurious.

Peaks with LOD-scores between 1.5 and 2.0 were detected on chromosomes 6, 7 and 10. Although those LOD-score were <2.0 and therefore linkage in these regions could be spurious, there is some overlap with other publications. The peaks on chromosomes 6 and 10 are (partly) overlapping with peaks for smoking initiation and for maximum number of cigarettes that we detected in a previous linkage analyses for in partly the same Dutch sample (Vink *et al.*, 2004). For chromosome 6, the results are a replication of a peak for smoking initiation with LOD score of 1.10 reported in the same region (63 cM) (Duggirala *et al.*, 1999). Although the regions on chromosomes 6 and 10 do not harbor obvious candidate genes for smoking, our results suggest that genes involved in smoking behavior may be located in this region.

Although age at onset of first cigarette is technically a continuous measure, in the present study it is collected as a quasi-continuous measure because answer categories were: 11 years or younger, 12, 13, 14, 15, 16, 17, 18 years or older and never tried. All subjects were aged 18 years or older. If the subjects who never tried smoking will start smoking in the future, they will change only one category (from “never” to “18 years or older”). In general it is accepted to analyze data with more than seven

categories as continuous variable (Dolan, 1994). An alternative is to use a threshold model where the trait is considered to have an underlying, continuous liability. However, using a threshold model reduces power (Neale *et al.*, 1994) and by combining categories information is lost.

It should be noted that the sample size in our study is relatively small and none of the linkage results met genome-wide statistical significance criteria. Due to the small sample size, there is a possibility for false-positive findings simply by chance. However, we determined the empirical *p*-value for chromosome 5, using a 1000× permutation algorithm. The empirical *p*-value of 0.04 for a LOD-score of 2.09 indicated that the peak on chromosome 5 (harboring the D1A dopamine receptor) is significant.

In conclusion, we detected 3 genomic regions with LOD score >2 (on chromosomes 5, 14 and 22) that may harbor susceptibility loci for age at first cigarette. A simultaneous three-locus analysis confirmed the importance of the peak on chromosome 5. There is some overlap between our results and those of previous genome scans looking at smoking behavior. To our knowledge, this is the first genome scan for age at smoking the first cigarette. The other genome scans to smoking behavior used different phenotypes. Each phenotype is likely to be influenced by multiple genes, leading to a different picture of the genetic architecture of smoking behavior.

ACKNOWLEDGMENTS

This work was supported by grants NWO 985-10-002 and NWO 904-61-090 from the Netherlands Organisation for Scientific Research (NWO) and grant 3100.0038 from ZonMW/NIDA and NIDA grant 018673. DP was supported by the GenomeEUtwin project (European Union Contract No. QLG2-CT-2002-01254). Genotyping was performed by the Marshfield Center for Medical Genetics. We would like to thank E. Suchiman and N. Lakenberg of the Department of Molecular Epidemiology in Leiden and J. Hottenga (CNCR) of the department of Biological Psychology in Amsterdam for their contributions.

REFERENCES

Abecasis, G. R., Cherny, S. S., Cookson, O. C., and Cardon, L. (2002). Merlin-rapid analyses of dense genetic maps using sparse gene flow trees. *Nature Genet.* **30**:97–101.

Abecasis, G. R., Cherny, S. S., Cookson, W. O., and Cardon, L. R. (2001). GRR: graphical representation of relationship errors. *Bioinformatics* **17**:742–743.

Bergen, A. W., Korszak, J. F., Weissbecker, K. A., and Goldstein, A. M. (1999). A genome-wide search for loci contributing to smoking and alcoholism. *Genet. Epidemiol.* **17**(S1):S55–S60.

Bierut, L. J., Rice, J. P., Goate, A., Hinrichs, A. L., Saccone, N. L., Foroud, T., Edenberg, H. J., Cloninger, C. R., Begleiter, H., Conneally, P. M., Crowe, R. R., and Hesselbrock, V. et al., (2004). A genomic scan for habitual smoking in families of alcoholics: common and specific genetic factors in substance dependence. *Am. J. Med. Genet.* **124**(A):19–27.

Boomsma, D. I., Vink, J. M., Beijsterveldt, C. E. M., Geus de, E. J. C., Beem, A. L., Mulder, E. J. C. M., Riese, H., Willemsen, A. H. M., Bartels, M., Berg van den, M., Derks, E. M., and Graaff, S. C. et al., (2002). Netherlands Twin Register: a focus on longitudinal research. *Twin Res.* **5**(5):401–406.

Comings, D. E., Gade, R., Wu, S., Chiu, C., Dietz, G., Muhleman, D., Saucier, G., Ferry, L., Rosenthal, R. J., Lesieur, H. R., Rugle, L. J., and MacMurray, P. (1997). Studies of the potential role of the dopamine D1 receptor gene in addictive behaviors. *Mol. Psychiat.* **2**(1):44–56.

Comuzzie, A. G., Blangero, J., Mahaney, M. C., Mitchell, B. M., Stern, M. P., and MacCluer, J. W. (1993). Quantitative genetics of sexual dimorphism in body fat measurements. *Am. J. Hum. Biol.* **6**:725–734.

Dolan, C. I. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: a comparison of categorical variable estimators using simulated data. *Brit. J. Math. Statist. Psychol.* **47**:309–326.

Duggirala, R., Almasy, L., and Blangero, J. (1999). Smoking behavior is under the influence of a major quantitative trait locus on human chromosome 5q. *Genet. Epidemiol.* **17**(S1):S139–S144.

Everret, S. A., Warren, C. W., Kann, S. D., Husten, C. G., and Crosset, L. S. (1999). Initiation of cigarette smoking and subsequent smoking behavior among U.S. high school students. *Prevent. Med.* **29**(5):327–333.

Fulker, D. W., and Cherny, S. S. (1996). An improved multipoint sib-pair analysis of quantitative traits. *Behav. Genet.* **26**(5):527–531.

Gelernter, J., Liu, X., Hesselbrock, V., Page, G. P., Goddard, A., and Zhang, G. (2004). Results of a genomewide linkage scan: support for chromosome 9 and 11 loci increasing risk for cigarette smoking. *Am. J. Med. Gen. Part B (Neuropsychiat. Genet.)* **128**:94–101.

de Geus, E. J. C., Posthuma, D., Kupper, N., Bergvan den, M., Willemsen, A. H. M., Beem, A. L., Slagboom, P. J., and Boomsma, D. I. (2005). A whole-genome scan for 24-hour respiration rate: a major locus at 10q26 influences respiration during sleep. *Am. J. Hum. Genet.* **76**:100–111.

Goring, H. H. H., Terwilliger, J. D., and Blangero, J. (2001). Large upward bias in estimation of locus-specific effects from genomewide scans. *Am. J. Hum. Genet.* **69**:1357–1369.

Heath, A. C., and Martin, N. C. (1993). Genetic models for the natural history of smoking: evidence for a genetic influence on smoking persistence. *Addict. Behav.* **18**:19–34.

Heath, A. C., Meyer, J. M., Kirk, K. M., and Martin, N. G. (1999). Genetic and social determinants of initiation of smoking in Australian twins. *Behav. Genet.* **29**(6):395–407.

Heijmans, B. T., Beekman, M., Putter, H., Lakenberg, N., van der Wijk, H. J., Whitfield, J. B., Posthuma, D., Pedersen, N., Martin, N. G., Boomsma, D. I., and Slagboom, P. E. (2005). Meta-analysis of four new genome scans for lipid parameters and analysis of positional candidates in positive linkage regions. *Eur. J. Hum. Genet.* **13**(10):1143–1153.

- Johnson, T. P., and Mott, J. A. (2001). The reliability of self-reported age of onset of tobacco, alcohol and illicit drug use. *Addiction* **96**:1187–1198.
- Kong, A., Gudbjartsson, D. F., Sainz, J., Jonsdottir, G. M., Gudjonsson, S. A., Richardson, B., Sigurdardottir, S., Barnard, J., Hallbeck, B., Masson, G., Shlien, A., and Palsson, S. T. et al., (2002). A high-resolution recombination map of the human genome. *Nature Genet.* **31**(3):241–247.
- Koopmans, J. R., Slutske, W., Heath, A. C., Neale, M. C., and Boomsma, D. I. (1999). The genetics of smoking initiation and quantity smoked in Dutch adolescent and young adult twins. *Behav. Genet.* **29**(6):383–393.
- Lamb, J. A., Barnby, G., Bonora, E., Sykes, N., Bacchelli, E., Blasi, F., Maestrini, E., Broxholme, J., Tzenova, J., Weeks, D. E., Bailey, A. J., and Monaco, A. P. (2005). Analysis of IMGSAC autism susceptibility loci: evidence for sex limited and parent of origin specific effects. *J. Med. Genet.* **42**(2):132–137.
- Lander, E. and Kruglyak, L. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genet.* **11**(November):241–247
- Lando, H. A., Thai, G. T., Murray, D. M., Robinson, L. A., Jeffery, R. W., Sherwood, N. E., and Hennrikus, D. J. (1999). Age of initiation, smoking patterns, and risk in a population of working adults. *Prevent. Med.* **29**(6):590–598.
- Lerman, C., Caporaso, N. E., Audrian, J., Main, D., Bowman, E. D., Lockshin, B., Boyd, N. R., and Shields, P. G. (1999). Evidence suggesting the role of specific genetic factors in cigarette smoking. *Health Psychol.* **18**(1):14–20.
- Li, M. D., Cheng, R., Ma, J. Z., and Swan, G. E. (2003a). A meta-analysis of estimated and environmental effects on smoking behavior in male and female adult twins. *Addiction* **98**(1):23–31.
- Li, M. D., Ma, J. Z., Cheng, R., Dupont, R. T., Williams, N. J., Crews, K. M., Payne, T. J., and Elston, R. C. (2003b). A genome-wide scan to identify loci for smoking rate in the Framingham Heart Study population. *BMC Genet.* **4**(Suppl 1):S103 .
- Ling, D., Niu, T., Feng, Y., Xing, H., and Xu, X. (2004). Association between polymorphism of the dopamine transporter gene and early smoking onset: an interaction risk on nicotine dependence. *J. Hum. Genet.* **49**:35–39.
- Lystig, T. C. (2003). Adjusted *P* values for genome-wide scans. *Genetics* **164**(4):1683–1687.
- Neale, M. C., Boker, S. M., Xie, G., and Maes, H. H. (1999). *Mx: statistical modeling*. Richmond, VA: Medical College of Virginia, Department of Psychiatry.
- Neale, M. C., Eaves, L. J., and Kendler, K. S. (1994). The power of the classical twin study to resolve variation in threshold traits. *Behav. Genet.* **24**(3):239–258.
- Posthuma, D., Luciano, M., de Geus, E. J. C., Wright, M., Slagboom, P. E., Montgomery, G. W., Boomsma, D. I., and Martin, N. G. (2005). A genomewide scan for intelligence identifies quantitative trait Loci on 2q and 6p. *Am. J. Hum. Genet.* **77**(2):318–326.
- Sabol, S. Z., Nelson, M. L., Fisher, G., Marcus, S. E., Gunzerath, L., Brody, C. L., Hu, S., Sirota, L. A., Greenberg, B. D., Lucas, F. R. IV, Benjamin, J., and Murphy, D. L. et al., (1999). A genetic association for cigarette smoking behavior. *Health Psychol.* **18**(1):7–13.
- Saccone, N. L., Neuman, R. J., Saccone, S. F., and Rice, J. P. (2003). Genetic analysis of maximum cigarette-use phenotypes. *BMC Genet.* **4**(Suppl 1):S105 .
- Sham, P. (1998). *Statistics in human genetics*. New York: Oxford University Press.
- Stallings, M. C., Hewitt, J. K., Beresford, T., Heath, A. C., and Eaves, L. J. (1999). A twin study of drinking and smoking onset and latencies from first use to regular use. *Behav. Genet.* **29**(6):409–421.
- Straub, R. E., Sullivan, P. F., Ma, Y., Myakishev, M. V., Harris-Kerr, C., Wormley, B., Kadami, B., Sadek, H., Silverman, M. A., Webb, B. T., Neale, M. C., and Bulik, C. M. et al., (1999). Susceptibility genes for nicotine dependence: a genome scan and followup in an independent sample suggest that regions on chromosomes 2, 4, 10, 16, 17 and 18 merit further study. *Mol. Psychiat.* **4**:129–144.
- Sullivan, P. F., and Kendler, K. S. (1999). The genetic epidemiology of smoking. *Nicotine Tobacco Res.* **1**:S51–S57.
- Sullivan, P. F., Neale, B., Oord van den, E., Miles, M. F., Neale, M. C., Bulik, C. M., Joyce, P. R., Straub, R. E., and Kendler, K. S. (2004). Candidate genes for nicotine dependence via linkage, epistasis, and bioinformatics. *Am. J. Med. Gen. Part B (Neuropsychiat. Genet.)* **126B**:23–36.
- Towne, B., Blangero, J., and Siervogel, R. M. (1993). Genotype by sex interaction in measures of lipids, lipoproteins, and apolipoproteins. *Genet. Epidemiol.* **10**:611–616.
- Towne, B., Siervogel, R. M., and Blangero, J. (1997). Effects of genotype-by-sex interaction on quantitative trait linkage analysis. *Genet. Epidemiol.* **14**:1053–1058.
- Vink, J. M., Beem, A. L., Posthuma, D., Neale, M. C., Willemsen, G., Kendler, K. S., Slagboom, P. E., and Boomsma, D. I. (2004). Linkage analysis of smoking initiation and quantity in Dutch sibling pairs. *Pharmacogenom. J.* **4**:274–282.
- Vink, J. M., Willemsen, G., and Boomsma, D. I. (2005). Heritability of smoking initiation and nicotine dependence. *Behav. Genet.* **35**(4):397–406.
- Wang, D., Ma, J. Z., and Li, M. D. (2005). Mapping and verification of susceptibility loci for smoking quantity using permutation linkage analysis. *Pharmacogenom. J.*: 1–7 (advance online publication).
- Weiss, L. A., Abney, M., Cook, E. H. J., and Ober, C. (2005). Sex-specific genetic architecture of whole blood serotonin levels. *Am. J. Hum. Genet.* **76**(1):33–41.
- Yuan, B., Vaske, D., Weber, J. L., Beck, J., and Sheffield, V. C. (1997). Improved set of short-tandem-repeat polymorphisms for screening the human genome. *Am. J. Hum. Genet.* **60**(2):459–460.