

# Genetic Time-series Analysis Identifies a Major QTL for *in vivo* Alcohol Metabolism not Predicted by *in vitro* Studies of Structural Protein Polymorphism at the *ADH1B* or *ADH1C* Loci

A. J. Birley,<sup>1,7</sup> J. B. Whitfield,<sup>1–3</sup> M. C. Neale,<sup>4</sup> D. L. Duffy,<sup>1</sup>  
A. C. Heath,<sup>5</sup> D. I. Boomsma,<sup>6</sup> and N. G. Martin<sup>1</sup>

Received 3 June 2004—Final 22 Feb. 2005

After ingestion of a standardized dose of ethanol, alcohol concentrations were assessed, over 3.5 hours from blood (six readings) and breath (10 readings) in a sample of 412 MZ and DZ twins who took part in an Alcohol Challenge Twin Study (ACTS). Nearly all participants were subsequently genotyped on two polymorphic SNPs in the *ADH1B* and *ADH1C* loci known to affect *in vitro* ADH activity. In the DZ pairs, 14 microsatellite markers covering a 20.5 cM region on chromosome 4 that includes the ADH gene family were assessed. Variation in the timed series of autocorrelated blood and breath alcohol readings was studied using a bivariate simplex design. The contribution of a quantitative trait locus (QTL) or QTL's linked to the ADH region was estimated via a mixture of likelihoods weighted by identity-by-descent probabilities. The effects of allelic substitution at the *ADH1B* and *ADH1C* loci were estimated in the means part of the model simultaneously with the effects sex and age. There was a major contribution to variance in alcohol metabolism due to a QTL which accounted for about 64% of the additive genetic covariation common to both blood and breath alcohol readings at the first time point. No effects of the *ADH1B*\*47His or *ADH1C*\*349Ile alleles on *in vivo* metabolism were observed, although these have been shown to have major effects *in vitro*. This implies that there is a major determinant of variation for *in vivo* alcohol metabolism in the ADH region that is not accounted for by these polymorphisms. Earlier analyses of these data suggested that alcohol metabolism is related to drinking behavior and imply that this QTL may be protective against alcohol dependence.

**KEY WORDS:** Alcohol metabolism *in vivo*; genetic time series; QTL.

<sup>1</sup> Queensland Institute of Medical Research and Joint Genetics Program, University of Queensland, 300 Herston Road, Herston, Brisbane, QLD, 4029, Australia.

<sup>2</sup> Department of Clinical Biochemistry, Royal Prince Alfred Hospital, Sydney, Australia.

<sup>3</sup> The University of Sydney, Australia.

<sup>4</sup> Virginia Institute of Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, Virginia.

<sup>5</sup> Missouri Alcoholism Research Centre, Department of Psychiatry, Washington University, St. Louis, Missouri.

<sup>6</sup> Department of Biological Psychology, Vrije Universiteit, Amsterdam, The Netherlands.

<sup>7</sup> To whom correspondence should be addressed at Queensland Institute of Medical Research, 300 Herston, Brisbane, QLD, 4029 Australia. Tel: +61-7-3362 0214; Fax: +61-7-3362 0101; e-mail: andrewBi@qimr.edu.au.

## INTRODUCTION

Alcohol metabolism shows considerable variation between individuals and this variation may have consequences for individual differences in duration of intoxication, alcohol intake, susceptibility to alcohol dependence and other alcohol-related traits and disorders. Because the initial and rate-limiting step in alcohol metabolism is its conversion to acetaldehyde by alcohol dehydrogenase (ADH), the alcohol dehydrogenase genes are prime candidates for the study of genetic variation in circulating alcohol concentrations after drinking.

The ADH genes consist of a family of seven genes at 4q22-4q23 (Osier, *et al.*, 2002; Tsukahara and Yoshida, 1989). One group of ADH genes, the Class I ADH genes, *ADH1A*, *ADH1B* and *ADH1C*, form a tandem, recently diverged and linked gene cluster (Yasunami, *et al.*, 1990). The transcribed products of the Class I genes interact to form the dimeric holoenzyme by way of subunit combination (Smith *et al.*, 1971, 1973). No polymorphisms that affect enzyme activity have been described for the *ADH1A* gene, but the effects of diallelic amino acid substitutions within each of the *ADH1B* and *ADH1C* genes upon *in vitro* enzyme kinetics are well documented (Bosron *et al.*, 1983; Edenberg and Bosron, 1997). Their kinetic properties led to an hypothesis for an etiological role in the development of problems associated with drinking (Chen *et al.*, 1999; Higuchi, *et al.*, 1995; Muramatsu, *et al.*, 1995; Osier *et al.*, 1999; Shen *et al.*, 1997; Whitfield *et al.*, 1997, 1998), by way of their role in the metabolism of ingested alcohol. Slower rates of alcohol metabolism enable the brain to experience the effects of alcohol for a longer time, which may lead to dependence. Conversely, when ethanol is metabolized quickly the disagreeable side effects of the first metabolite, acetaldehyde, can become apparent. The most compelling evidence for this effect involves the *ADH1B* locus in populations of East Asian descent (Thomasson *et al.*, 1991), where the gene frequency of the *ADH1B\*2* allele (high *in vitro* enzyme activity), reaches a frequency of circa 60%. This is not the case in Caucasians where the *ADH1B\*2* allele has a population frequency of about 10%. Protection through avoidance of alcohol is probably provided by the deterrent effects of accumulated acetaldehyde. Such an effect is particularly noticeable in those who, in addition to carrying the *ADH1B\*2* allele, also carry the inactive acetaldehyde dehydrogenase (ALDH) allele *ALDH2\*2* (see references in Chen *et al.*, 1999; Osier *et al.*, 1999), but the effects of ALDH2 deficiency and of the *ADH1B\*2* allele are independently significant. There is also evidence for an association of *ADH1B* with alcohol use and dependence in Caucasians (Borras, *et al.*, 2000; Whitfield *et al.*, 1998).

The most plausible biochemical and physiological explanation of these population based associations was thought to be found in the considerable differences in the *in vitro*,  $V_{max}$ , (the maximum rate of ethanol oxidation when all ADH enzyme is complexed with ethanol) of proteins encoded by the *ADH1B* and *ADH1C* loci. Since the combined difference in the *in vitro*  $V_{max}$  of the allozymes produced

by the two most extreme homozygotes at the *ADH1B* and *ADH1C* loci is about 40-fold (Edenberg and Bosron, 1997), an evaluation of the contribution of these genes to *in vivo* alcohol metabolism variation is of particular interest.

We evaluated the contributions of *ADH1B* and *ADH1C* to *in vivo* metabolism using a joint association – linkage approach. The simultaneous analysis of association and linkage makes it possible to identify other variants in addition to *ADH1B* and *ADH1C* in the linked ADH gene cluster which may influence alcohol metabolism. These variants may include sequences concerned with transcription and regulation which can be distant from their target gene or gene family member (Blackwood and Kadonaga, 1998; Engel, 1993; Orkin, 1990). Furthermore, whilst ingested alcohol may not be the ancestral substrate for all ADH genes, apart from *ADH5*, they all may act on ethanol at high concentration (Beisswenger, *et al.*, 1985; Edenberg, 2000), and with the further exception of *ADH6* and *ADH7*, are expressed in the human liver and are major contributors to ethanol detoxification (Zgombic-Knight, *et al.*, 1995). Although allelic polymorphisms at the *ADH1B* and *ADH1C* loci undoubtedly cause variation in the *in vitro* kinetic properties, their contributions to the overall genetic variance in alcohol metabolism, *in vivo*, have yet to be assessed in a family design. Interestingly the *ADH1B\*2* allele accounted for only a small proportion (8.5%) of the total phenotypic variance of *in vivo* metabolism in a Jewish population in which it reaches a frequency of 21% (Neumark *et al.*, 2004). In this same study there was no evidence for a contribution from the *ADH1C* locus.

A study of *in vivo* alcohol metabolism in Australian twins and known as the Alcohol Challenge Twin Study (ACTS), was carried out over the period, 1979–1981 (Martin *et al.*, 1985). It has provided a unique resource as a set of time-related profiles of *in vivo* alcohol metabolism in young adult MZ and DZ twins. Heritabilities of the peak blood alcohol level and rate of alcohol elimination were found to be high ( $0.62 \pm 0.06$  and  $0.50 \pm 0.11$  respectively) in this sample.

In this paper we use multivariate structural equation modeling (MSEM) of the blood and breath readings of alcohol concentrations to model the effect of a potential QTL (Boomsma and Dolan, 1998, 2000; Eaves *et al.*, 1996; Martin *et al.*, 1997). MSEM was used to analyze the variation in time-dependent

trajectories of alcohol level as a function of IBD status in DZ twin pairs. The model implementation was done in the program Mx (Neale, 1997), which allows the simultaneous modeling of covariance and means structures by the method of full information maximum likelihood. The model for the means allows the contribution of biologically important covariates such as age and sex, as well as the effects of measured genotypes at the *ADH1B* and *ADH1C* loci to be estimated. The inclusion of these covariates increases the chance of satisfying the assumption of multivariate normality of residuals implicit in the method, which is especially important for genes likely to be of large effect. It also provides a way of estimating the contribution of polymorphisms at the *ADH1B* and *ADH1C* loci to the QTL variance arising from the ADH region as a whole.

Several theoretical studies have shown that a multivariate analysis of correlated traits can result in an increase in power to detect a QTL as compared to individual univariate measures (Boomsma and Dolan, 1998; Martin *et al.*, 1997; Schmitz *et al.*, 1998). A simplex model (Boomsma and Molenaar, 1987; Eaves *et al.*, 1986) was used to specify the covariance structure for timed reading of alcohol concentration. This quasi-Markovian model is more suited to the analysis of the time-dependent auto-correlated data, than is univariate analysis as it models the effects of time. It is also unconstrained by any particular kinetic model, which is reasonable given the complexity of alcohol metabolism, transport and exchange within the body.

Two related hypotheses of the genetic causes of *in vivo* alcohol metabolism are tested by reduced versions of the model: (1) there is a QTL within the ADH region that accounts for a significant fraction of the genetic variance associated with alcohol metabolism *in vivo*; and (2) alleles at the *ADH1B* and *ADH1C* loci are major contributors to this QTL, as would be predicted from differences in their *in vitro*  $V_{max}$ .

## MATERIALS AND METHODS

### The Subjects and the Measurement of Alcohol Metabolism

The 206 twin pairs who were participants in the ACTS 1979–81, (Martin *et al.*, 1985) were of Caucasian (mainly Northern European) ancestry. They comprise 88 MZ (45 female and 43 male) twin pairs and 118 DZ (42 female, 37 male and 39 opposite sex) twin pairs. The age range of our subjects was quite narrow (18–34) with 70% of subjects < age 25.

Measures of blood and breath alcohol levels (expressed as mg 100 ml<sup>-1</sup>) were recorded for twins following ingestion over 20 minutes of a weight-related dose of ethanol (0.75 g kg<sup>-1</sup>) (Martin *et al.*, 1985). The measures of breath alcohol were made equivalent to those of blood alcohol by multiplying by the blood/breath factor of 2100:1 (Whitfield and Martin, 1994) and both measures followed a Gaussian distribution. The measurement apparatus was sometimes unreliable and both this and a revision of the protocol during this study (Whitfield and Martin, 1994) led to some missing values (Table I), particularly for breath alcohol readings.

Blood samples were not retained after the ACTS was completed. By 1990 it was apparent that genotyping of individual subjects from whom data on blood alcohol levels after alcohol challenge was already available would allow testing of genotype–phenotype associations. Blood was collected from 372 of the original 412 subjects and typed for the *ADH1B\*47His* and *ADH1C\*349Ile* polymorphisms (Whitfield *et al.*, 1998). DNA from these subjects was used to obtain genotypes for a set of 14 microsatellite markers that spanned the *ADH* gene region.

### Genetic Markers

Of the 14 microsatellite markers, three – D4S231 and D4S2634 (centromeric) and *EGF* (telomeric)-flanked the cytogenetic region 4q22–4q25 that contains the *ADH* gene family. The remaining 11 – D4S2380, D4S1628,

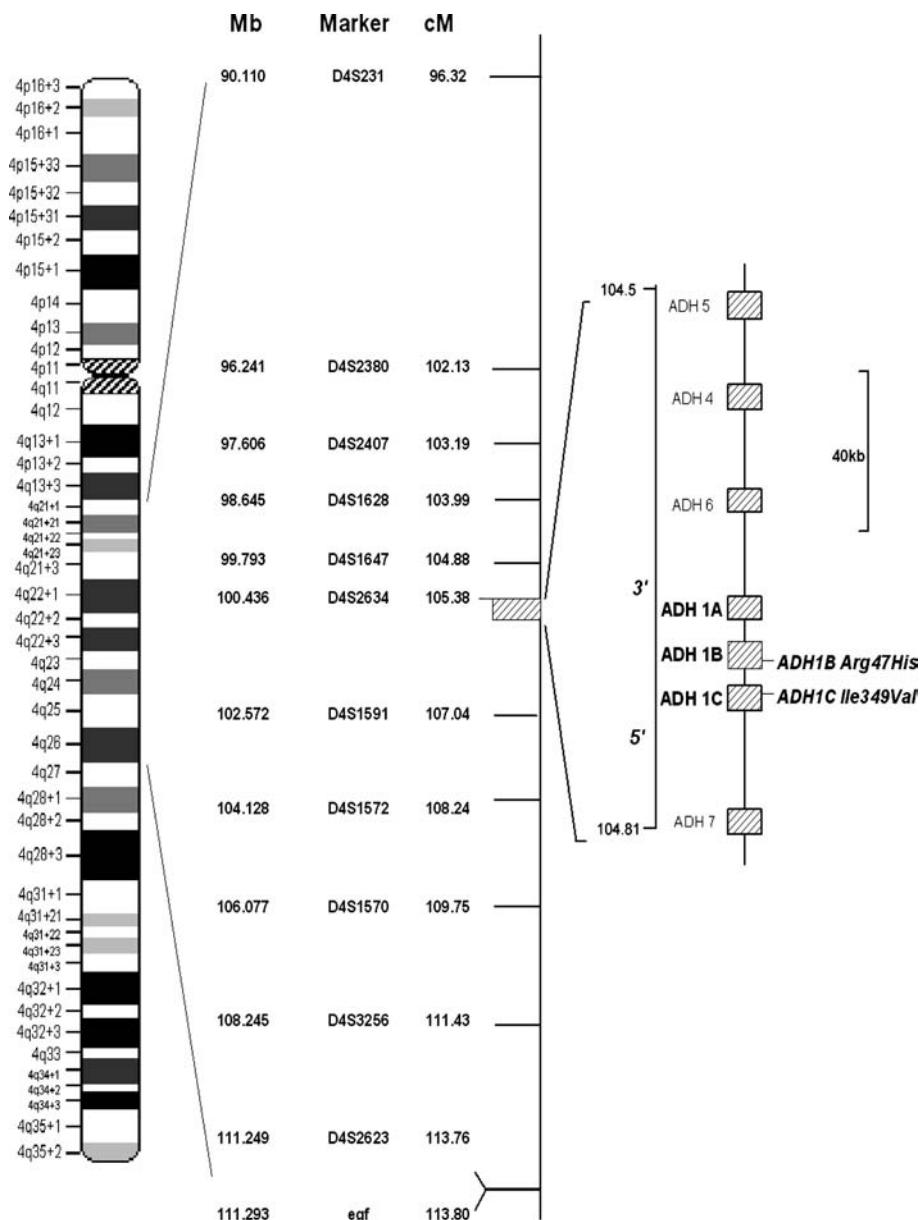
**Table I.** Number of Twin Subjects for Whom Blood and Breath Alcohol Levels were Recorded in Relation to Zygosity

Time after ingestion (minute)	Zygosity	40	56	68	83	100	123	143	160	182	213
Blood alcohol	MZ	–	165	172	164	–	176	102	–	171	–
	DZ	–	223	235	223	–	235	167	–	232	–
Breath alcohol	MZ	168	161	168	158	168	167	94	168	168	168
	DZ	226	221	232	222	232	231	163	232	231	232

D4S1647, D4S2407, D4S1591, D4S1572, D4S1570, D4S2623, D4S3256, and *ADH1B* (formerly, *ADH2*) (Duester *et al.*, 1999) and *ADH1C* (formerly, *ADH3*) were tightly enclosed within the ADH gene family. Polymorphism at the *ADH1B* locus is identified by the alleles *ADH1B\*47His* (formerly *ADH1B\*2*) and *ADH1B\*47Arg* (formerly *ADH1B\*1*). Polymorphism at *ADH1C* is identified as *ADH1C\*349Ile*, (formerly *ADH1C\*1*) and *ADH1C\*349Val* (formerly *ADH1C\*2*).

Both loci were identified by PCR/RFLP methods (Whitfield *et al.*, 1998). For brevity we have referred to the alleles at these polymorphic loci by the simpler notation *ADH1B\*1* or *ADH1C\*1* etc. A total of 105 DZ and 78 MZ twin pairs were so genotyped at the *ADH1B* and *ADH1C* (respectively *ADH1B\*Arg47His* and *ADH1C\*Ile349Val*) loci.

The locations and map-order of the markers and organization of the ADH gene cluster on the



**Fig. 1.** Recombination map locations of genetic markers and location of the ADH gene family. The overall map is presented in physical order and the direction of transcription of the *ADH1* gene cluster is given.

physical map (Fig. 1) was obtained from the UCSC Genome Browser website; (Karolchik *et al.*, 2003). The DeCODE recombination map (Kong *et al.*, 2002) was accessed with 'Cartographer' software (National Public Health Institute Helsinki, Finland website) and provided a genetic map position for all markers except D4S231, *ADH1B*, *ADH1C* and *EGF*. Nucleotide map positions of all markers were obtained from the UCSC Genome Browser web site with the 'BLAT Search' option. Primer assay sequences for the *ADH1B*\*Arg47His and *ADH1C*\*Ile349Val loci were identified from within snp\_view tables located within the NCBI website, and locations cited within Osier *et al.* (1999). The polymorphic site D4S231 was located indirectly by taking the average location in Mb of the respective and proximal and distal WI-18650 and B337WH9 markers within the UCSC Genome Browser. The tetranucleotide microsatellite marker of *EGF* is located within intron 10 (Murray *et al.*, 1992). A sex-averaged linear calibration of 1 MB = 0.775046 cM for markers located on the DeCODE map enabled unknown genetic distances to be relationally placed on the recombination map.

IBD status was estimated for DZ twin pairs using the above map. DNA was also obtained from one or more parents for 63 of the 105 DZ twin pairs and/or from one or more non-twin siblings of 55 of the DZ pairs. These additional family members contributed to checks on mistyping and to the estimates of IBD status. Spurious double recombinants likely to be the result of mistyping were identified by detailed scanning of the IBD probabilities over the 25–30 cM interval.

The probabilities of IBD sharing status of 0, 1 or 2 alleles ( $p_0$ ,  $p_1$ ,  $p_2$ ) were calculated over the ADH region for DZ pairs using the multipoint method of Kruglyak and Lander (1995) implemented in MAPMAKERSIBS2.1. For convenience we used IBD probabilities estimated at a point intermediate in position between *ADH1B* and *ADH1C* although the calculated IBD probabilities were very well polarized and consistent across the genotyped region. Haplotype configurations for *ADH1B* and *ADH1C* were determined using SNPHAP (SNPHAP website). A posterior probability of 0.85 was used to screen unresolved haplotypes.

#### Structural Equation Modeling and Approach to QTL Detection

Maximum likelihood methods were used to estimate the contribution of genes and environment

to the time course of alcohol metabolism. They were implemented with the structural equation modeling package Mx (Neale, 1997) version 1.54 using the raw data input option. The option enables missing data points to be accommodated within the analysis following the procedure of Lange *et al.* (1976). All analyses were carried out as mg. ethanol.100 ml.<sup>-1</sup>, divided by 10 to facilitate numerical optimization in Mx.

The approach to QTL detection involved the parameterization of the expected twin covariance matrices conditional upon the pairs sharing  $k$  alleles IBD. The method maximizes the log-likelihood of the observations and the marker data at particular locations across the genome. Following Eaves *et al.* (1996) and Neale (2000), we define the log-likelihood ( $\ln L_1$ ) of a sibling pair, sharing  $k$  alleles IBD (assuming that the distribution of the phenotypes, on  $k$  follows a multivariate normal distribution), as:

$$\ln \mathbf{L}(\mathbf{X}_i, \mathbf{k}) = \ln |\Sigma_k|^{-n} - \mathbf{0.5}(\mathbf{x}_i - \boldsymbol{\mu})' \Sigma_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu})$$

where  $\Sigma_k$  is the  $(2n \times 2n)$  expected covariance matrix for  $n$  variables measured on the  $i$ th sibling pair and  $|\cdot|$  is the determinant of  $\Sigma_k$ . The means of the phenotypes are given by the column vector  $\boldsymbol{\mu}$ . The difference between this log-likelihood (or  $\ln L_1$ ) and the log-likelihood,  $\ln L_0$  in which a selected subset of  $j$  parameters are fixed at hypothesized parameter values, provides the basis of the log-likelihood ratio test statistic,  $T = 2(\ln L_1 - \ln L_0)$ , which is asymptotically distributed as a  $\chi^2$  value for  $j$  degrees of freedom. In the present model, hypothesis testing falls into two categories, one concerned with the parameters modeled in the expected sib pair covariance matrix and a second for the tests of hypotheses of the predicted means vector for the  $i$ th sib pair.

Previous analyses (Martin *et al.*, 1985) and preliminary multivariate analyses did not support the inclusion of shared family environment effects in the expected covariance model which was parameterized by the three covariance matrices: the residual or background genetic matrix  $\mathbf{A}$ , within family environment matrix  $\mathbf{E}$ , and the matrix containing the contributions of the linked QTL effects,  $\mathbf{Q}$ .

$$\begin{pmatrix} \mathbf{A} + \mathbf{E} + \mathbf{Q} & 0.5\mathbf{A} + \pi_k \mathbf{Q} \\ 0.5\mathbf{A} + \pi_k \mathbf{Q} & \mathbf{A} + \mathbf{E} + \mathbf{Q} \end{pmatrix},$$

where  $\pi_k$ , is the expected proportion of a random sample of sib pairs sib pairs taken from a panmictic population, that share  $k$  alleles IBD where  $k = (0, 1, \text{ or } 2)$ .

The weighted likelihood method (Dolan *et al.*, 1999; Fulker and Cherny, 1996; Neale, 2000) was used to estimate QTL effects. An expected covariance matrix for a DZ pair is defined for each of the three IBD states 0, 1 and 2 in which  $\pi_k = k/2$ . The probabilities for the  $i$ th DZ pair for the three IBD states of 0, 1 and 2 were used as the mixing proportions of the three expected covariance matrices. The expected panmictic probabilities of 0.25, 0.5 and 0.25, corresponding to the respective IBD states 0, 1 and 2, were used (Eaves *et al.*, 1996) for those 7 DZ twin pairs for whom information on IBD status was unavailable. The assignment of the expected panmictic IBD probabilities to ungenotyped sibling pairs has been criticized (Schork and Greenwood, 2004) because it leads to bias towards the null hypothesis with some methods of linkage detection (Cordell, 2004). As shown by Dolan *et al.*, (1999), when calculated IBD's are available only for phenotypes at the extremes of the phenotypic distribution the pi-hat method of QTL estimation leads to a marked upward bias in the test statistic; this is not the case when the weighted likelihood model is used. As noted by Cordell (2004), the methods we have adopted in this study for the treatment of missing IBD information should not result in bias. Although no information about linkage is obtained from MZ twins, they were included in the analysis as a separate group since they are informative in the estimation of the effects of the measured genotypes at the *ADH1B* and *ADH1C* loci and in the estimation of the total heritability of the traits. Cholesky decomposition was used to saturate the covariance structure due to specific environmental effects, since these were of secondary interest and this ensures conservative tests of the genetic components of covariance.

### Bivariate Time-series Modeling of Blood and Breath Alcohol Levels

The simplex model provides a natural description of covariation from repeated measurements of the same subjects across time (Boomsma and Molenaar, 1987; Eaves, *et al.*, 1986) such as our series of timed measurements of blood and breath alcohol levels that describe the effects of *in vivo* alcohol metabolism following the prescribed dose of ethanol. The bivariate simplex process (Boomsma and Molenaar, 1987) was used to model the QTL (*Q*) and the residual genetic (*A*) effects common to both blood and breath measures as well as QTL and residual genetic effects specific to either blood or breath measurements. The bivariate simplex model (Fig. 3)

was used to estimate the time dependent innovations and the transmission parameters of the common and the specific time series. Details of how the model was specified and of the procedure used to estimate the latent residual genetic (or QTL) parameters are given in the Appendix A.

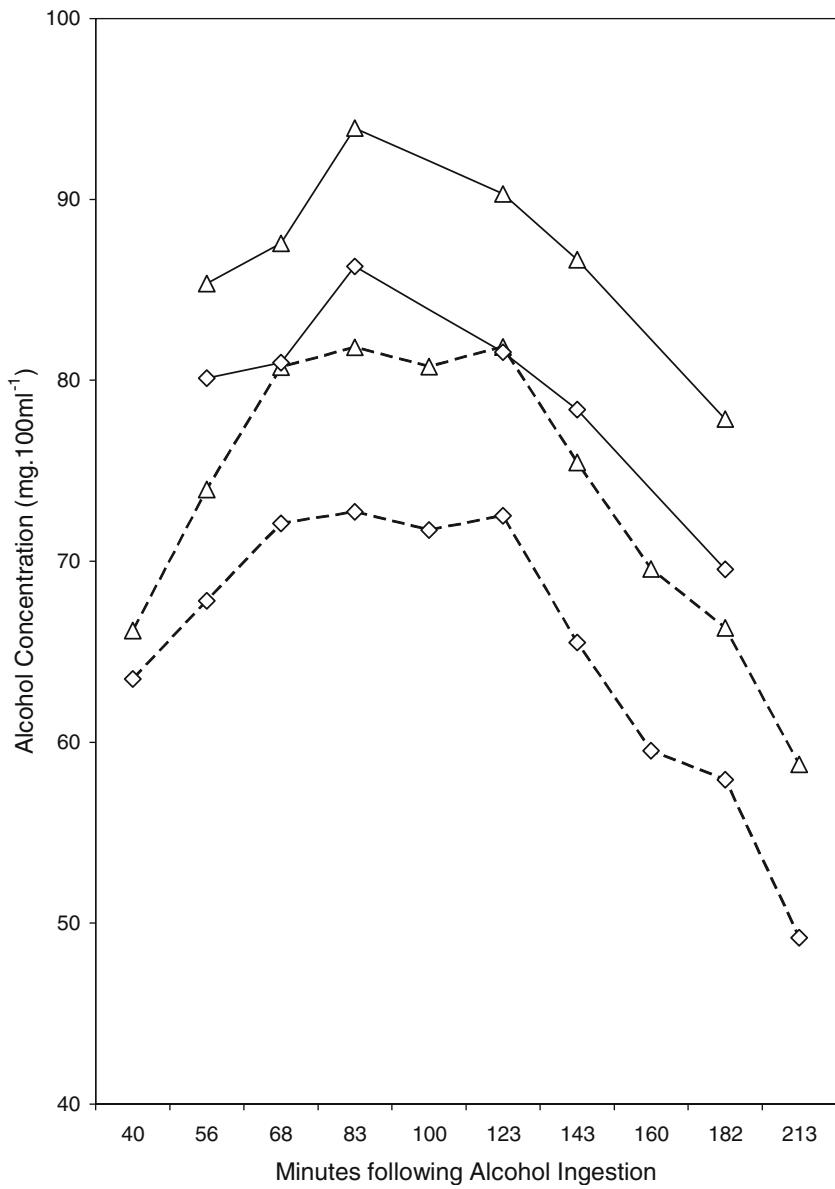
### The Means Model

The predicted means vector was defined for each time point as a grand mean,  $\mathbf{m}$ , and a matrix of regression weights  $\boldsymbol{\beta}_k$ , post multiplied by a vector of  $k_{ij}$  covariate values, for the  $j$ th sibling of the  $i$ th sibpair and  $k$ th covariate. Maximum likelihood estimates (MLE) of means effects, sex and age and the effects of haplotypes at the *ADH1B* and *ADH1C* loci are obtained jointly with the QTL and residual genetic covariances. The method due to Fulker *et al.* (1999) provides a joint test the contribution of the measured genotypes to linkage. The amino acid substitutions that identify the *ADH1B*\* Arg47His and *ADH1C*\*Ile349Val loci lead to marked changes of *in vitro*  $V_{max}$  so making them good candidates for *in vivo* ADH metabolism. If the effects of the measured genotypes are singularly causal to the observed linkage they will absorb a significant proportion of the QTL variation.

## RESULTS

### The Effects of Sex and Age

Males had lower blood and breath alcohol concentrations (BAC, Fig. 2) than females. For breath readings this effect was strongly associated with time and was greatest towards the end of the time series ( $\chi^2_9 = 34.12$ ;  $p < 0.0001$ ). A similar trend was apparent but not statistically significant for blood readings ( $\chi^2_5 = 4.52$ ;  $p = 0.48$ ) which, compared to the readings for breath alcohol level, covered a more restricted time span that centered at the height of the parabolic relation for the time course of alcohol metabolism. The effect of sex measured as a male deviation amounted to an average of  $-9.13 \text{ mg } 100 \text{ ml}^{-1}$  for breath and  $-8.82 \text{ mg } 100 \text{ ml}^{-1}$  for blood alcohol concentration. There was no age  $\times$  time interaction effect on either blood ( $\chi^2_5 = 4.57$ ;  $p = 0.47$ ) or breath ( $\chi^2_9 = 14.79$ ;  $p = 0.10$ ) alcohol concentration level. There was a clear linear regression of alcohol level upon age, (as a time independent effect), of  $0.04 \text{ mg } 100 \text{ ml}^{-1} \text{ yr}^{-1}$  for both blood and breath measurements ( $\chi^2_1 = 9.04$ ;  $p < 0.001$  and  $\chi^2_1 = 12.66$ ;



**Fig. 2.** Time course of Blood (—) and Breath alcohol (- - -) concentration (mg ml<sup>-1</sup>) after for Female (△) and Male (◊) participants in the Alcohol Twin Challenge Study.

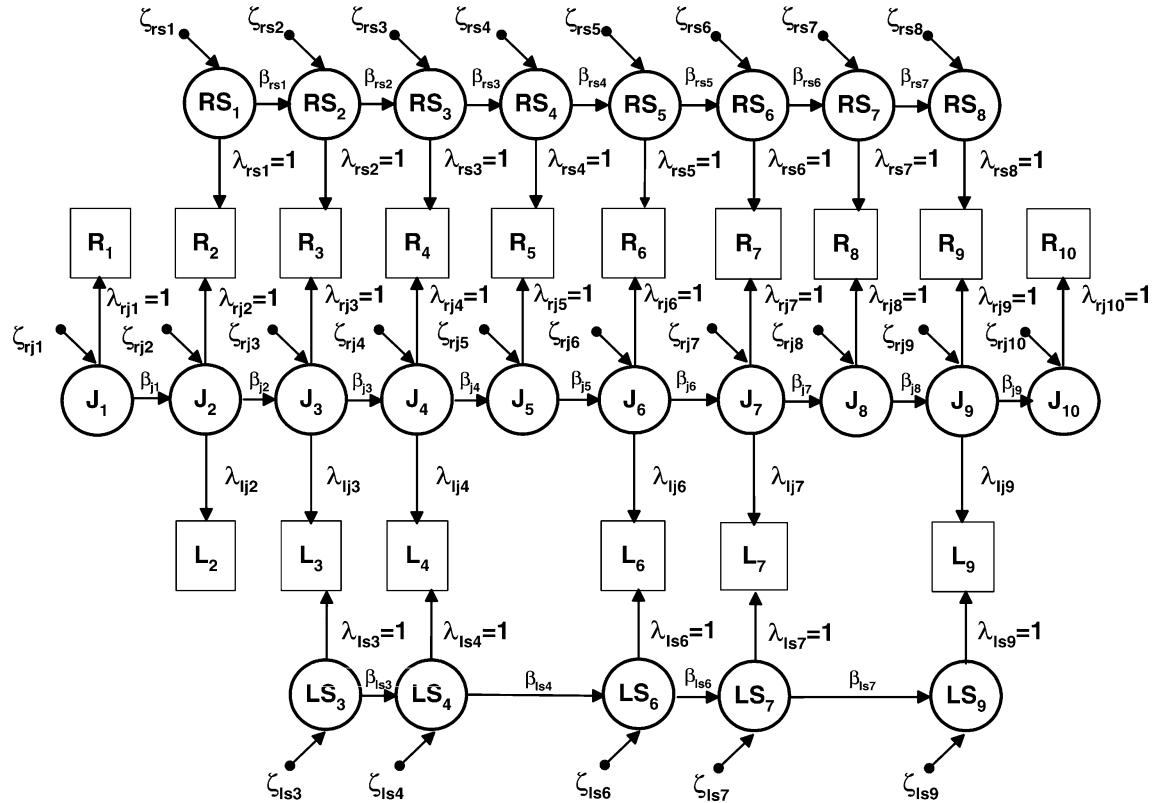
$p < 0.001$  respectively. Both sex and age were retained in the means model for all subsequent modeling of covariance structure.

### Genetic Modeling

#### *QTL Detection*

The contribution of a putative QTL to the expected sib pair covariance matrices involved the specification of an expected covariance matrix  $\mathbf{Q}$ . The

effects of  $Q$  were modeled through a bivariate (i.e. blood and breath measurements) simplex model that was identical to that used for the residual genetic covariance matrix  $A$ , except for the genetic correlations between genotypes of twins. These correlations depend on IBD status for  $Q$  and were fixed at 0.5 for  $A$  for DZ twins. Identification of the bivariate simplex models for  $A$  and  $\mathbf{Q}$  was achieved by starting the simplex processes that are specific to blood or specific to breath at the second time of measurement



**Fig. 3.** Specification of the bivariate simplex model for the time course of the observed measurements  $\square$ . There are 10 time points for breath and 6 time points for the blood alcohol concentrations. The joint latent genetic ( $A$ ) (or QTL ( $Q$ )) process as estimated from the covariance between blood and breath alcohol readings is indicated by  $\circ$  as with labels  $J_{1-10}$ . The symbols,  $RS_{2-8}$  or  $LS_{2,3,4,6,7}$  and  $9$ , relate to those simplex processes that are respectively specific to breath or blood measures at the indicated times. Innovations of variance ( $\zeta_{rs1-rj10}$ ) are shown for the common genetic (or QTL) covariance, the covariance specific to blood and that specific breath, are shown as  $\bullet$ . They are depicted in the identified model in relation to the latent variables. Transmission paths ( $\beta$ ) for the specific process for blood, breath and those for the process common to both breath and blood are related to the timing of readings by the respective suffixes, ( $\beta_{ls}$ ,  $\beta_{rs}$ , and  $\beta_j$ ). The transmission paths for blood readings take into account missing time points. Latent or  $\lambda$  variables are specified for the common pathway to enable the relative contributions of blood and breath readings in the time series to be compared. The model shown in this diagram is fitted for both  $A$  and  $Q$ . In addition a Cholesky decomposition of  $E$  is fitted.

(Boomsma, *et al.*, 1989). The structure of the model describing the QTL structure was simplified from the full bivariate simplex structure, whilst the model for the residual genetic covariance was retained to ensure conservative tests of QTL effects. A set of nested sub models of the full weighted likelihood model for QTL effects provided tests of the contribution of the QTL to both the specific (blood or breath measurements) and to the common simplex processes (Table II).

Simplification of the fully specified bivariate simplex QTL model proceeded by testing for the contribution of the QTL processes specific to blood, and that specific to breath. Both specific processes could be dropped from the simplex structural model that defined the QTL (Table II, Model 2); the genetic components of the variation specific to either blood

or breath levels are unrelated to IBD status. In striking contrast, dropping the entire part of the simplex model concerned with the QTL for the common covariance ( $Q_J$ ) between blood and breath measures from the model (Table II, Model 3) led to a drastic worsening of model fit ( $\chi^2_{19} = 65.99$ ;  $p < 10^{-8}$ ).

The first innovation variance of the QTL process is unique in our simplified model for covariance associated with the QTL (Fig. 4) and inspection of the relative magnitude of the first innovation for  $Q_J$  (Fig. 4) indicated that most genetic variance due to the QTL does indeed arise at the beginning of the time series. On dropping the first QTL innovation in the time series (Table II, Model 4), the change of fit was considerable ( $\chi^2_1 = 34.41$ ;  $p < 10^{-8}$ ), showing

**Table II.** Simplification of the Bivariate Simplex Model (Fig. 3) for QTL Effects

Model	Comparison <sup>a</sup>	Residual genetic covariance <sup>b</sup>			QTL covariance <sup>c</sup>			Goodness of fit				
		$A_J$	$A_{RS}$	$A_{LS}$	$Q_J$	$Q_{RS}$	$Q_{LS}$	-2LL	df	$\chi^2$	df	
1	1	Y	Y	Y	Y	Y	Y	12963.18	5812			
2	1	Y	Y	Y	Y	N	N	12980.68	5836	17.492	24	0.83
3	2	Y	Y	Y	N	N	N	13046.67	5855	65.995	19	$<10^{-6}$
4	2	Y	Y	Y	$Y^d$	N	N	13015.09	5837	34.412	1	$<10^{-8}$

<sup>a</sup>The baseline model for tests of the change in log likelihood and hence model fit.

<sup>b</sup> $A_J$ ,  $A_{RS}$  and  $A_{LS}$  are those partitions of the model for the latent residual genetic covariance that respectively describe the joint covariance breath and blood measures that which is specific to breath, and that process specific to blood.

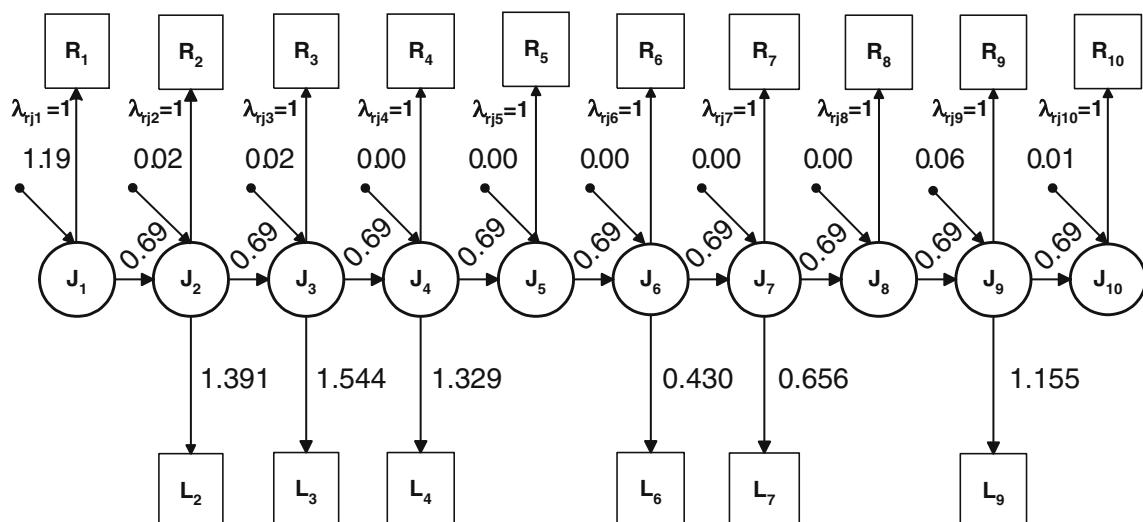
<sup>c</sup> $Q_J$ ,  $Q_{RS}$  and  $Q_{LS}$  are the corresponding effects described above for the contribution of the linked QTL.

that the QTL for alcohol metabolism is indeed manifested soon after drinking. The model for  $Q_J$  was found to further simplify from that defined by Model 2. A reduced form of QTL structure (Fig. 4) was obtained by equating the nine transmission paths for  $Q_J$  ( $\chi^2_9 = 13.38$ ;  $p = 0.10$ ), where the relative size of the innovation of variance at time 1 can be seen relative to the contributions at subsequent time points. Indeed dropping the nine small innovations of covariance subsequent to the major innovation seen at time 1 did not result in a reduction in model fit ( $\chi^2_9 = 15.044$ ;  $p = 0.09$ ). The reduced model for  $Q_J$  led to a change of fit of ( $\chi^2_2 = 37.57$ ;  $p < 10^{-8}$ ), which is very similar to Model 4 in (Table II). The percentage of total genetic variance in alcohol metabolism accounted for by the QTL specified by the innovation

at the first time point was 64% with 95% CI's of 56.7% and 66.5% (estimated by the method of Neale and Miller (1997)).

#### Effects of Haplotypes at the *ADH1B* and *ADH1C* Loci

The effects of haplotypes (Table III) at the biallelic *ADH1B* and *ADH1C* loci were estimated in the means part of the expected likelihood equation. Three of the four possible haplotypes namely *ADH1C\*1-ADH1B\*1* (52.8%), *ADH1C\*2-ADH1B\*1* (43.1%), and *ADH1C\*1 - ADH1B\*2* (4.1%) were identified (SNPHAP website) in this sibpair sample. The population frequencies of the *ADH1C\*1* and *ADH1B\*2* alleles in the ACTS sample were 0.569 and 0.037 respectively.



**Fig. 4.** Model for QTL covariance showing the contributions of blood and breath measures to the QTL for alcohol metabolism following the nomenclature given in Fig. 3. The joint latent QTL process ( $Q_{J1} - Q_{J10}$ ) between blood and breath alcohol readings is indicated here by  $J_1$  to  $J_{10}$ .

**Table III.** Haplotypes at the *ADH1C* and *ADH1B* Loci and their Effects upon Alcohol Metabolism

Haplotype <sup>a</sup>	1-1	1-2	2-1
Frequency	0.528	0.431	0.041
Effect <sup>b</sup>	0.45	0.51	3.22
% QTL variance <sup>c</sup>	<0.01%	<0.01%	0.90%

<sup>a</sup>Three haplotypes *ADH1C\*1-ADH1B\*1* (1-1), *ADH1C\*1-ADH1B\*2* (1-2), and *ADH1C\*2-ADH1B\*1* (2-1).

<sup>b</sup>Effect size ( $\text{mg.100 ml}^{-1}$ ) estimated in the means part of the ML equation and equated over all measurement times.

<sup>c</sup>Percentage change of QTL variance accounted for the effects of a haplotype.

Phase was only a potential problem for a small part of the sample; i.e. for those cases (15) for which both loci are heterozygous and all of these were at  $p > 0.85$ .

The standardized coefficient of linkage disequilibrium  $D'$  value (or  $D/D_{\max}$ , Lewontin, 1964, 1988) for this sample was  $-0.76$ . There was no evidence of heterogeneity of haplotype effects in relation to times at which the blood or breath alcohol readings were taken. Similarly the effects of the haplotypes, respectively encoded (in 5'-3' orientation) as 1-1, 2-1 and 1-2, upon blood or breath alcohol readings were homogeneous when constrained across time. The effect size (measured as deviation from the mean) of the 1-1 haplotype was  $0.45 \text{ mg } 100 \text{ ml}^{-1}$  ( $\chi_1^2 = 0.26$ ;  $p = 0.61$ ), for the 2-1 haplotype it was  $0.53 \text{ mg } 100 \text{ ml}^{-1}$  ( $\chi_1^2 = 2.7$ ;  $p = 0.10$ ), whilst that for the 1-2 haplotype was  $3.26 \text{ mg } 100 \text{ ml}^{-1}$  ( $\chi_1^2 = 3.84$ ;  $p = 0.05$ ). The contribution of the 1-2 haplotype, although small is associated with the effects of the *ADH1C\*1* and *ADH1B\*2* alleles and represents that combination of alleles with the highest *in vitro* activities at these loci. The association due to haplotypes provides the most informative test of the hypothesis that our QTL is due to polymorphism at the *ADH1C* and *ADH1B* loci and was marginally more sensitive than one based upon allelic association alone ( $\chi_1^2 = 2.32$ ;  $p = 0.12$  for the *ADH1B* locus and  $\chi_1^2 = 2.35$ ;  $p = 0.13$  for the *ADH1C* locus). However, although the effect of the 1-2 haplotype is significantly different from zero ( $p = 0.05$ ), its effect size is small and its contribution to the genetic variance attributable to the first QTL innovation of covariance common to both breath and blood measures was only 0.9%. Given the low frequency of *ADH1B\*2*, the contribution of this locus to the genetic variance can be expected to be small in this Caucasian population and account for a negligible fraction of the genetic variance due to the QTL for alcohol metabolism.

## DISCUSSION

Phenotypic assessment of ethanol metabolism was done in 206 pairs of twins by measuring blood and BAC over a three hour time course following ingestion of a weight related dose of ethanol. The 105 DZ twins for whom we had DNA were typed for 14 markers in the ADH region of chromosome 4. Sib pair linkage analysis was used to see if we could detect a linked QTL that accounted for a significant proportion of the variance in alcohol metabolism. Sib pair linkage has notoriously low power (Risch and Zhang, 1995). We began by performing linkage analysis of the 16 individual blood and breath QTL readings and in no case found a drop chisquare of  $> 4$  (i.e. LOD  $< 1$ ) at the ADH region. The point estimates of percentage variance attributable to a QTL and 95% CI's (in brackets) for the six successive measurements of blood alcohol concentrations were 0 (0–0.411), 0.197 (0–0.625), 0 (0–0.608), 0.121 (0–0.532), 0.46 (0–0.778) and 0.502 (0–0.719) and similarly for the 10 breath alcohol times they were 0 (0–0.443), 0 (0–0.276), 0 (0–0.338), 0 (0–0.462), 0 (0–0.211), 0 (0–0.265), 0 (0–0.366), 0.154 (0–0.549), 0 (0–0.369) and 0.173 (0–0.538). Whilst most estimates of the QTL are poor, the upper 95% CI's are often greater than 50% of the phenotypic variance. Results like these are discussed by Martin *et al.* (1997) where the power to detect a QTL effect was found to be demonstrably greater (over five times) for a multivariate approach as compared to analysis of the mean phenotype. In our case the inconsistency of estimates is partly due variability in the degree of completeness of the data. In that case the multivariate approach enables all of the information to be incorporated in a joint analysis. Indeed the proportion of phenotypic variance due a QTL and its 95%CI for the mean of the standardized scores of blood measures was 0.001 (0–0.0531) for breath measures, 0 (0–0.265) and of both blood and breath measures, 0 (0–0.367). Schmitz *et al.* (1998), demonstrated that multivariate genetic analysis can lead to substantial increases in the power to detect genetic variation by making use of the correlations between timed measurements such that even the present small sample sizes can be realistically utilized when heritabilities are greater than about 0.5 and correlations between time points are in the range 0.3–0.6 (Schmitz *et al.*, 1998). The pattern of phenotypic correlations declined over time and corresponded to those that identify a simplex rather than a common factor model. Moreover multivariate modeling of QTL effects is especially effective in Mx

where data points are missing for some time points and/or phenotypes (Table I), as a univariate analysis of every time point and phenotype would be based upon incomplete data sets that differ in their form and extent between time points, measures of alcohol concentration.

We used a bivariate simplex analysis to jointly estimate the effects of a QTL on all 10 breath and 6 blood alcohol readings as a parallel time series. This multivariate analysis has dramatically increased our power to detect a QTL. Our analysis suggests that 64% of genetic variance in alcohol metabolism is due to a QTL in the ADH region. The inclusion of fixed effects due to haplotypes at the polymorphic *ADH1B* and *ADH1C* loci suggests that only a negligible proportion of the QTL variance is attributable to these polymorphisms.

One intriguing aspect of QTL detection concerns the negative correlation between estimates of the QTL contribution and those of the residual genetic variance. The effect of this relationship, applicable to both univariate and multivariate analysis, is to reduce the power to detect a QTL, a factor that is dependent upon the degree to which the residual genetic variance is solely due to stochastic events. These effects may increase during a time series. Our estimates of QTL effects and their statistical significance and confidence limits from the joint simplex model in the light of this relationship are the most optimistic that we are likely to achieve. The most conservative approach to QTL detection with these data is a partition of the residual genetic covariance structure in the unsimplified form of full Cholesky decomposition. This led to a model of QTL structure for the common process of the bivariate simplex with only a major innovation at time point 1. The chi-square value for 1 df obtained by dropping this innovation of variance from the model was 12.54 ( $p < 0.001$ ).

The simplex model provides a natural description of changes of covariance structure over time that result from the physiology of alcohol absorption and elimination, part of which is due to the *in vivo* enzyme kinetics of alcohol oxidation. It is particularly useful in the present study, as a physiological model that satisfactorily describes the dynamics of blood alcohol absorption detoxification and elimination from the body has yet to appear. Moreover, the proportion of variation accounted for by this QTL will necessitate its inclusion into future general models of alcohol metabolism. Genetic variation for alcohol metabolism is innovated immediately after ingestion of alcohol. As such, the model is capable of detecting

gradual changes in effects of the same or different genes that may be either regulatory or structural in nature and which are realized as innovations of variance in the time series.

Haplotypes for the *ADH1B* and *ADH1C* loci showed very little evidence of anything but a minor contribution to genetic variation in *vitro* metabolism although alleles at both loci are known to be widely different their *in vitro*  $V_{max}$  (Bosron, *et al.*, 1983). Although the haplotype *ADH1B\*2-ADH1C\*1* increases the metabolism of ethanol it is at low frequency (4.1%), and its effect was at borderline significance. Allele *ADH1B\*2* is certainly at low frequency in this Caucasian sample and can therefore make, at most, a small contribution to the overall genetic variation in these phenotypes. Given the low frequency of the *ADH1B\*47His* allele, the locus will also contribute little to linkage in general. In contrast, the QTL that we have identified by linkage to the genomic region containing the ADH gene family accounts for a considerable fraction of the genetic variance for alcohol metabolism.

The joint assessment of QTL linkage and allelic association (Fulker *et al.*, 1999) also allows for the detection of population admixture as a cause of allelic association. Given the combination of small effect sizes and statistical significance for tests of association involving the *ADH1B* and *ADH1C* loci, any such further partition would be uninformative. A stronger association of *ADH1C* locus with alcohol consumption in the same ACTS subjects was not due to population stratification (Neale *et al.*, 1999). The ability to the contribution of a candidate locus to the effect linkage in our model is also dependent upon genetic architecture. If the candidate locus is itself the causal to an observed QTL effect then the model described by Fulker *et al.* (1999) provides a powerful approach to the cause of an observe linkage and will be particularly effective if the common disease – common variant view of genetic architecture is the case (Almasy and Blangero, 2004). Nonetheless the power to detect the effect of association will be poor if there the gene(s) of interest are tightly linked to the test locus but not in LD with the test locus. This is certainly a possibility since LD is not always directly related to map distance. Although in principle it may be more difficult to detect the contribution of many rare variants segregating within families to linkage (should our genetic architecture conform to the rare variant – common “disease” model), newer approaches in haplotype based methods suggest a more optimistic outlook (Lin *et al.*, 2004). The application

of extensive SNP genotyping to genes within ADH gene region will permit further tests of the contribution of those polymorphisms that are at high frequency to linkage.

We already know of other genes within the ADH gene family that may have a significant role in ethanol oxidation, but their contributions in the context of our QTL are still to be investigated. For example, the class II gene, *ADH4* ( $\pi$ -ADH), is mainly expressed in the human liver. It has a relatively high  $K_m$  for ethanol and at intoxicating levels of alcohol it may account for up to 40% of ethanol oxidation in the liver (Li *et al.*, 1977). A polymorphic promoter site polymorphism is also associated with *ADH4* that accounts for a twofold difference in *ADH4* expression *in vitro* (Edenberg *et al.*, 1999). *ADH6* (Class V) is also expressed at its highest levels in the liver (Zhi *et al.*, 2000). *ADH7* (Class IV;  $\sigma$ -ADH) has a high  $V_{max}$  and high  $K_m$  for ethanol and NAD<sup>+</sup>, suggesting it is effective at high alcohol level, and is mainly expressed in the stomach mucosa and oesophagus; it has also been implicated as being protective against alcoholism (Osier *et al.*, 2004). The wider possibility of transcriptional regulation within the ADH region also needs to be considered with the possibility of gene-control sequences capable of influencing an effect over long tracts of DNA (Groupp *et al.*, 1994). Interestingly, the retention of gene family clusters in evolution is related to the frequency and duration of transcription by way of a single locus control region (Grosveld, *et al.*, 1987; Milot *et al.*, 1996) but an evolutionary history of gene duplication does not necessarily lead to any particular ordered logic in the coding sequences of gene families or even of their transcriptional control. Although the *ADH4* locus may be a candidate, the extent of non-coding sequences involved in transcriptional regulation of the ADH region has yet to be determined.

The relationship between the *ADH1B* and *ADH1C* genes and the risk of alcohol dependence and other drinking habits, is thought to arise from the effects of rapid alcohol metabolism in the presence of the *ADH1B*\*47His allele in people of East Asian descent. There is now increasing evidence for an involvement of the ADH region in alcohol dependence in populations of different ethnicity. The COGA study provides strong evidence for linkage within the ADH region in a genome scan for DSMIIIIR diagnosis of alcohol dependence (Reich *et al.*, 1998) and for DSM-IV diagnosis (Williams *et al.*, 1999). Chromosome 4 linkage close to the ADH region for alcohol dependence is also supported in

extended pedigrees from a South Western American tribe in which the *ADH1B*\*2 allele is absent (Long *et al.*, 1998). An indicative relationship between alcohol dependence (DSM IIIIR) and blood and breath alcohols has been demonstrated in the ACTS twins (Whitfield *et al.*, 2001).

Linkage disequilibrium (LD) across the ADH gene cluster (Osier *et al.*, 2002) makes it difficult to identify causal relationships between polymorphic alleles and phenotypes concerned with drinking habits. The *ADH1C* (alleles *ADH1C*\*1 and *ADH1C*\*2) and *ADH1B* (alleles *ADH1B*\*1 and *ADH1B*\*2) genes and their intergenic sequences span about 15 kb (Yasunami, *et al.*, 1990), and show strong LD with a  $D'$  value (or  $D/D_{max}$ , (Lewontin, 1964, 1988)) of  $\approx 0.7$  that is no different in alcoholics as compared to non-alcoholics (Osier *et al.*, 1999). There is a gross disparity in allele frequencies for the *ADH1B*\*2 allele between East Asian and Caucasian ethnic groups and haplotype analysis has also revealed that the genetic architecture surrounding the *ADH1B*\*2 allele differs between the two ethnic groups (Osier *et al.*, 2002). LD between the *ADH1B* and *ADH1C* loci in the ACTS population was also estimated from genotype frequencies using the EH and UNKNOWN algorithms (Terwilliger and Ott, 1994) so as to include information from both DZ and MZ twins and other family members. The standardized coefficient of linkage disequilibrium  $D'$  was  $-0.75$ , close to that seen in East Asian populations; a good agreement given the allele frequency divergence between Caucasian and East Asian peoples. Hence any geographic or ethnic differences in the risk for alcoholism associated with variation at the *ADH1B* locus may also be due to differences in the background genotype, and/or cultural differences.

The physiological consequences of a QTL accounting for 64% of the genetic variation for alcohol metabolism are likely to be manifold. The QTL will contribute to the variance in neuropharmacological effects in the brain because of its effect on alcohol level following drinking, and by way of the broad substrate specificity of the Class I and II enzymes. Elevated alcohol levels increase neurotransmitter function in a variety of ways; notable effects are increased dopamine, serotonin, endogenous opioid and GABA receptor function and a decrease in glutamate receptor function. Recent evaluations of the kinetics of class I and II enzymes show that the high NADH/NAD<sup>+</sup> ratio produced during alcohol detoxification is also favorable for the reduction of aldehydes to alcohols. For example, following deamination of serotonin (5-hydroxytrypt-

amine) by monamine oxidase, the aldehyde product is normally oxidized by aldehyde oxidase to hydroxyindoleacetic acid (5HIAA). However, following drinking, aldehyde dehydrogenase is fully saturated by acetaldehyde from alcohol oxidation by ADH (Helander *et al.*, 1993), and the raised NADH/NAD<sup>+</sup> ratio favors the production of 5-hydroxytryptophol (5HTOL) from 5HIAA by ADH (Svensson *et al.*, 1999). Indeed after alcohol ingestion the ratio of 5HTOL/5HIAA increases in a dose related manner (Helander *et al.*, 1994). This concomitant increase in serotonin metabolism may be a factor in the development of hangover and depression related symptoms. A notable association on chromosome 15q for single nucleotide polymorphisms in the *GABRG3* gene with alcohol dependence illustrates how the addiction may come about in individuals with very excited nervous systems (Dick *et al.*, 2004). Clearly the physiological and genetic effects of the QTL for alcohol metabolism identified here are beginning to unfold. Our task is now to fine map this region and identify the variants responsible for this huge QTL effect on ethanol metabolism.

## APPENDIX A

Following the notation of Boomsma and Molenaar (1987) the residual additive genetic covariance  $\Sigma_A$ , can be generally formulated for the simplex design as:

$$\Sigma_A = \Lambda_y(\mathbf{I} - \mathbf{B})^{-1}\psi(\mathbf{I} - \mathbf{B}')^{-1}\Lambda'_y. \quad (1)$$

Time related latent factor loadings ( $\lambda_1$  through  $\lambda_t$ ) of the observations on the genetic factors are estimated in the lambda matrix,  $\Lambda_y$ . In this form, the latent genetic covariance structure ( $g$ ) depends upon both the magnitude of new innovations of genetic variance estimated at times 1 ...  $t$ , and estimated in a diagonal matrix  $\psi$  (by parameters  $\zeta_{g1}$  through  $\zeta_{gt}$ ), and the proportion of genetic variance ( $\beta_{g1}$  through  $\beta_{gr-1}$ ) transmitted to a total of ( $t - 1$ ) succeeding time points. These latter parameters are referred to as transmission paths and are estimated from the sub-diagonal elements of the  $\beta$  matrix. The structure of these matrices are explicitly shown in Boomsma and Molenaar (1987).

The residual genetic variance  $A$ , at any one time point  $t$ , can be calculated as

$$A_t = \lambda_t^2(\beta_t^2 A_{t-1}) + \varsigma_t. \quad (2)$$

When presented in this form, both  $\lambda_t^2$  and  $\varsigma_t^2$  provide the same estimate of variance at time  $t$ , (Boomsma, *et al.*, 1989) and as such equation (2) provides two alternative approaches to estimation. On the one hand the variance of the innovations  $\varsigma_t$ , can be set at unity, in which case the variance of the latent genetic factor at time  $t$  becomes:

$$\lambda_t^2(\beta_t^2 A_{t-1}) + 1. \quad (3)$$

Alternatively, the latent factor loadings can be fixed at unity such that the estimate of the latent genetic factor at time  $t$  is provided by the time dependent innovation  $\varsigma_t$ .

$$(\beta_t^2 A_{t-1}) + \varsigma_t. \quad (4)$$

The present bivariate simplex analysis of the blood and breath measurements is based upon that presented by Boomsma *et al.* (1989) and used both approaches to estimation without compromising the identification of the model.

The application of the bivariate simplex model to the timed series of blood and breath measurements involves a specification of the  $\Lambda$ ,  $B$  and  $\psi$  matrices to allow for the partition of effects that are specific to the blood measures as well as those that are specific to the breath measures from the common genetic covariance between the blood and breath measures for the series of timed measurements, conditional upon readings for both phenotypes being available at the same time.

The general specification of the matrices shown in equation (1), then forms the basis of the estimation procedure for the bivariate simplex model. The  $\lambda$  matrix in its bivariate form (size 20,30) contains six sub-matrices in its structure:

$$\begin{bmatrix} \mathbf{C} & \mathbf{D} & \mathbf{O} \\ \mathbf{G} & \mathbf{O} & \mathbf{F} \end{bmatrix},$$

where the diagonal matrices  $\mathbf{C}$  and  $\mathbf{G}$  (each size 10,10) contain estimates of the latent genetic factors that correspond to 10 breath and 6 blood measurements). On the other hand diagonal matrices  $\mathbf{D}$  and  $\mathbf{F}$  enable the estimation of those latent genetic factors specific to breath or blood. A zero matrix  $\mathbf{O}$  (size 10,10) completes the necessary substructure of the lambda matrix. A matrix size (30,30) with a substructure of nine matrices each size 10,10 defines the  $\mathbf{B}$  matrix of transmission paths as:

$$\begin{bmatrix} \mathbf{L} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{M} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{N} \end{bmatrix},$$

The transmission paths are estimated as sub diagonal elements of matrices  $\mathbf{L}$ ,  $\mathbf{M}$  and  $\mathbf{N}$ ; those specified in Matrix  $\mathbf{L}$  are estimated in relation to the common covariance for breath and blood, whilst matrices  $\mathbf{M}$  and  $\mathbf{N}$  correspond to the transmission paths that are respectively specific to breath and specific to blood. In this part of the model, for the specific effects only 8 (breath) and 4 (blood) transmission paths are identified. The  $\psi$  matrix (size 30,30) is structured to allow groups of 30 parameters or innovation variances to be estimated on the leading diagonal of an otherwise zero matrix. The groups each size 10 are represented in the  $\psi$  matrix in the order – common genetic effects, specific breath and blood specific effects. For identification the simplex specific processes specific to the either blood or breath readings were parameterized such that neither the innovation at the first time point in the  $\psi$  matrix nor the first transmission path in the time series concerned with either of the specific processes were estimated (Fig. 3). As illustrated by equation (2) there are two possible ways of ways in which a simplex model can be parameterised. Here the series of innovation variances in the  $\psi$  matrix the contributions of the two (blood and breath) specific genetic time series were specified such that their corresponding  $\lambda$  elements in the submatrices  $\mathbf{D}$  and  $\mathbf{F}$  were set to unity.

This led to the estimation of a time related series of innovations of variance ( $\zeta$  parameters) and their corresponding transmission paths (as  $\beta$  parameters), for those parts of the simplex model that measured effects specific to blood ( $\zeta_{ls3} - \zeta_{ls5}$ ) and ( $\beta_{ls3} - \beta_{ls7}$ ) or to breath ( $\zeta_{rs2} - \zeta_{rs8}$ ) and ( $\beta_{rs2} - \beta_{rs7}$ ) (see Fig. 3). The contributions of the measurements from blood and those from breath to their common genetic covariance were modeled by setting the genetic factor loadings on the joint process (estimated in submatrix  $\mathbf{C}$  of the  $\psi$  matrix) for breath to unity. In contrast the contributions of the blood measures to the common process are seen indirectly by estimating their genetic factor loadings (parameters  $\lambda_{lj2} - \lambda_{lj9}$ ) for the common process in submatrix  $\mathbf{G}$  matrix, whilst the corresponding elements in the  $\psi$  matrix were set to unity (Fig. 3). The estimates ( $\lambda_{lj2} - \lambda_{lj9}$ ) are of the relative contributions of breath measures to blood contribu-

tions to the common genetic covariance on those six occasions for which both blood and breath measures were available. The common genetic covariance was modeled as a timed series of innovations,  $\zeta_{rj1} - \zeta_{rj10}$  and transmission paths,  $\beta_{j1}$  to  $\beta_{j9}$ , respectively as described above (Fig. 3).

## ELECTRONIC-DATABASE INFORMATION

The URLs for data presented herein are as follows:

- UCSC Genome Browser website (<http://genome.ucsc.edu/> release hg16; NCBI Build 34).
- Cartographer software (<http://www.bioinfo.helsinki.fi/cartographer>).
- NCBI website – `snp_view` tables, [http://www.ncbi.nlm.nih.gov/SNP/snp\\_viewTable.cgi?h=KIDDLAB](http://www.ncbi.nlm.nih.gov/SNP/snp_viewTable.cgi?h=KIDDLAB).
- SNPHAP, <http://www-gene.cimr.cam.ac.uk/clayton/software/snphap.txt>.

## ACKNOWLEDGEMENTS

We should like to acknowledge the contributions of our erstwhile collaborators in collecting the phenotype data, particularly John Oakeshott, John Gibson, Graham Starmer and Judy Perl. We thank Connor Dolan for his assistance with the longitudinal data analysis. We should also like to thank the twins and their families for continuing collaboration in providing DNA samples. We are grateful to Barbara Haddon for coordination of DNA sample collection and Carol Mayne, Brian Nightingale and Martin O'Brien for expert technical assistance in genotyping. This work was funded by grants from the Australian Brewers Foundation, NHMRC (#159101), NIAAA (AA077281, AA013321, AA013326, AA014041) and NIMH (MH-65322).

## REFERENCES

- Almasy, L., and Blangero, J. (2004). Exploring positional candidate genes: linkage conditional on measured genotype. *Behav. Genet.* **34**:173–179.
- Beisswenger, T. B., Holmquist, B., and Vallee, B. L. (1985). ChiADH is the sole alcohol dehydrogenase isozyme of mammalian brains: implications and inferences. *Proc. Natl. Acad. Sci. USA* **82**:8369–8373.
- Blackwood, E. M., and Kadonaga, J. T. (1998). Going the distance: a current view of Enhancer action. *Science* **281**:60–63.
- Boomsma, D. I., and Molenaar, P. C. M. (1987). The genetic analysis of repeated measures. I. Simplex models. *Behav. Genet.* **17**:111–122.

- Boomsma, D. I., Martin, N. G., and Molenaar, P. C. M. (1989). Factor and simplex models for repeated measures: application to psychomotor measures of alcohol sensitivity in twins. *Behav. Genet.* **19**:79–96.
- Boomsma, D. I., and Dolan, C. V. (1998). A comparison of power to detect a QTL in sib-pair data using multivariate phenotypes, mean phenotypes, and factor-scores. *Behav. Genet.* **28**:329–340.
- Boomsma, D. I., and Dolan, C. V. (2000). Multivariate QTL analysis using structural equation modeling: a look at power under simple conditions. In H. Spector, A. Snieder and MacGregor (eds.), *Advances in Twin and Sib-Pair Analysis*. Greenwich Medical Media Ltd , pp. 203–218.
- Borras, E., Coutelle, C., Rosell, A., Fernandez-Muixí, F., Broch, M., Crosas, B., Hjelmqvist, L., Lorenzo, A., Gutierrez, C., Santos, M., Szczepanek, M., Heilig, M., Quattrochi, P., Farres, J., Vidal, F., Richard, C., Mach, T., Bogdal, J., Jörnvall, H., Seitz, H. K., Couzgou, P., and Pares, X. (2000). Genetic polymorphisms of alcohol dehydrogenase in Europeans: the ADH2\*2 allele decreases the risk for alcoholism and is associated with ADH3\*1. *Hepatology* **31**:984–989.
- Bosron, W. F., Magnes, L. J., and Li, T.-K. (1983). Kinetic and electrophoretic properties of native and recombinant isoenzymes of human liver alcohol dehydrogenase. *Biochemistry* **22**:1852–1857.
- Chen, H. J., Carr, K., Jerome, R. E., and Edenberg, H. J. (2002). A retroviral repetitive element confers tissue-specificity to the human alcohol dehydrogenase 1C (ADH1C) gene. *DNA Cell Biol.* **21**:793–801.
- Cordell, H. J. (2004). Bias toward the null hypothesis in model-free linkage analysis is highly dependent on the test statistic used. *Am. J. Hum. Genet.* **74**:1294–1302.
- Dick, D. M., Edenberg, H. J., Xuei, X., Goate, A., Kuperman, S., Schuckit, M., Crowe, R., Smith, T. L., Porjesz, B., Begleiter, H., and Foroud, T. (2004). Association of GABRG3 with alcohol dependence. *Alcohol Clin. Exp. Res.* **28**:4–9.
- Dolan, C. V., Boomsma, D. I., and Neale, M. C. (1999). A simulation study of the effects of assignment of prior identity-by-descent probabilities to unselected sib pairs, in covariance-structure modeling of a quantitative-trait locus. *Am. J. Hum. Genet.* **64**:268–280.
- Duester, G., Farrés, J., Felder, M. R., Holmes, R. S., Höög, J.-O., Parés, X., Plapp, B. V., Yin, S. J., and Jörnvall, H. (1999). Recommended nomenclature for the vertebrate alcohol dehydrogenase gene family. *Biochem. Pharmacol.* **58**:389–395.
- Eaves, L. J., Long, J., and Heath, A. C. (1986). A theory of developmental change in quantitative phenotypes applied to cognitive development. *Behav. Genet.* **16**:143–162.
- Eaves, L. J., Neale, M. C., and Maes, H. (1996). Multivariate multipoint linkage analysis of quantitative trait loci. *Behav. Genet.* **26**:519–525.
- Edenberg, H. J. (2000). Regulation of the mammalian alcohol dehydrogenase genes. *Prog. Nucleic Acid Res. Mol. Biol.* **64**:295–341.
- Edenberg, H. J., and Bosron, W. F. (1997). Alcohol dehydrogenase. In: F. P. Guengreich (ed.), *Biotransformation*. I. G., McQueen C. A., Gandolfi, A. J. (eds.), *Comprehensive Toxicology Sipes*. New York: Pergamon, pp. 119–131.
- Edenberg, H. J., Jerome, R. E., and Li, M. (1999). Polymorphism of the human alcohol dehydrogenase 4 ADH4 promoter affects gene expression. *Pharmacogenetics* **9**:25–30.
- Engel, J. D. (1993). Developmental regulation of human beta-globin gene transcription: a switch of loyalties?. *Trends Genet.* **9**:304–309.
- Fulker, D. W., and Cherny, S. S. (1996). An improved multipoint sibpair analysis of quantitative traits. *Behav. Genet.* **26**:527–532.
- Fulker, D. W., Cherny, S. S., Sham, P. C., and Hewitt, J. K. (1999). Combined linkage and association sib pair analysis for quantitative traits. *Am. J. Hum. Genet.* **64**:259–267.
- Grosveld, F., Assendelft, G. B. van, Greaves, D. R., and Kollias, G. (1987). Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* **51**:975–985.
- Groupp, E. R., Crawford, N., and Locker, J. (1994). Characterization of the distal alpha-fetoprotein enhancer, a strong, long distance, liver-specific activator. *J. Biol. Chem.* **269**:22178–22187.
- Helander, A., Beck, O., Jacobsson, G., Löwenmo, C., Wikström, T. (1993). Time course of ethanol-induced changes in serotonin metabolism. *Life Sci.* **53**:847–855.
- Helander, A., Beck, O., and Borg, S. (1994). The use of 5-hydroxytryptophol as an alcohol intake marker. *Alcohol Alcoholism (Suppl.)* **2**:497–502.
- Higuchi, S., Matsushita, S., Murayama, M., Takagi, S., and Hayashida, M. (1995). Alcohol and aldehyde dehydrogenase polymorphisms and the risk for alcoholism. *Am. J. Psychiatry* **152**:1219–1221.
- Karolchik, D., Baertsch, R., Diekhans, M., Furey, T. S., Hinrichs, A., Lu, Y. T., Roskin, K. M., Schwartz, M., Sugnet, C. W., Thomas, J., Weber, R. J., Haussler, D., and Kent, W. J. (2003). The UCSC genome browser database. *Nucl. Acids Res.* **31**:51–54.
- Kong, A., Gudbjartsson, D. F., Sainz, J., Jónsdóttir, G. M., Gudjonsson, S. A., Richardson, B., Sigurdardóttir, S., Barnard, J., Hallbeck, B., Masson, G., Shlién, A., Palsson, S. T., Frigge, M. L., Thorleifsson, T. E., Gulcher, J. R., and Stefansson, K. (2002). A high-resolution recombination map of the human genome. *Nat. Genet.* **31**:225–226.
- Kruglyak, L., and Lander, E. S. (1995). Complete multipoint sibpair analysis of qualitative and quantitative traits. *Am. J. Hum. Genet.* **57**:439–454.
- Lange, K., Westlake, J., and Spence, M. (1976). Extensions to pedigree analysis: III. Variance components by the scoring method. *Ann. Hum. Genet.* **39**:485–491.
- Lewontin, R. C. (1964). The interaction of selection and linkage. I. General considerations: heterotic models. *Genetics* **49**:49–67.
- Lewontin, R. C. (1988). On measures of gametic disequilibrium. *Genetics* **120**:849–852.
- Long, J. C., Knowler, W. C., Hanson, R. L., Robin, R. W., Urbanek, M., Moore, E., Bennett, P. H., and Goldman, D. (1998). Evidence for genetic linkage to alcohol dependence on chromosomes 4 and 11 from an autosome-wide scan in an American Indian population. *Am. J. Med. Genet.* **81**:216–221.
- Li, T.-K., Bosron, W. F., Dafeldecker, W. P., Lange, L. G., and Vallee, B. L. (1977). Isolation of PI-alcohol dehydrogenase of human liver: is it a determinant of alcoholism?. *Proc. Natl. Acad. Sci.* **74**:4378–4381.
- Lin, S., Chakravarti, A., and Cutler, D. J. (2004). Exhaustive allelic transmission disequilibrium tests as a new approach to genome-wide association studies. *Nat. Genet.* **36**:1181–1188.
- Martin, N. G., Perl, J., Oakeshott, J. G., Gibson, J. B., Starmer, G. A., and Wilks, A. V. (1985). A twin study of ethanol metabolism. *Behav. Genet.* **15**:103–109.
- Martin, N. G., Boomsma, D. I., and Machin, G. (1997). A twin-pronged attack on complex traits. *Nat. Genet.* **17**:387–392.
- Milot, E., Strouboulis, J., Trimborn, T., Wijgerde, M., Boer, E. de, Langeveld, A., Tan-Un, K., Vergeer, W., Yannoutsos, N., Grosveld, F., and Fraser, P. (1996). Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell* **87**:105–114.
- Muramatsu, T., Wang, Z. C., Fang, Y. R., Hu, K. B., Yan, H., Yamada, K., Higuchi, S., Harada, S., and Kono, H. (1995). Alcohol and aldehyde dehydrogenase genotypes and drinking behavior of Chinese living in Shanghai. *Hum. Genet.* **96**:151–154.
- Murray, J. C., Bennett, S. R., Kwitek, A. E., Small, K. W., Schinzel, A., Alward, W. L., Weber, J. L., Bell, G. I., and Buetow, K. H. (1992). Linkage of Rieger syndrome to the region of the epidermal growth factor gene on chromosome 4. *Nat. Genet.* **2**:46–49.

- Neale, M.C. (1997). *Mx: Statistical Modeling*. 3rd ed Richmond, VA: Department of Psychiatry, Medical College of Virginia, Virginia Commonwealth University
- Neale, M. C. (2000). QTL mapping with sib-pairs: the flexibility of Mx. In T. D. Spector, H. Snieder and A. J. MacGregor (eds.), *Advances in Twin and Sib-pair Analysis*. London: Greenwich Medical Media Ltd., pp. 219–243.
- Neale, M. C., and Miller, M. B. (1997). The use of likelihood based confidence intervals in genetic models. *Behav. Genet.* **27**:113–120.
- Neale, M. C., Cherny, S. S., Sham, P. C., Whitfield, J. B., Heath, A. C., Birley, A. J., and Martin, N. G. (1999). Distinguishing population stratification from genuine allelic effects with Mx: Association of ADH1B with alcohol consumption. *Behav. Genet.* **29**:233–243.
- Neumark, Y. D., Friedlander, Y., Durst, R., Leitersdorf, D. J., Ramchandani, V. A., O'Connor, S., Carr, L. G., and Li, T.-K. (2004). Alcohol dehydrogenase polymorphisms influence alcohol-elimination rates in a male Jewish population. *Alcohol Clin. Exp. Res.* **28**:1014.
- Orkin, S. H. (1990). Globin gene regulation and switching: circa 1990. *Cell* **63**:665–672.
- Osier, M., Pakstis, A. J., Kidd, J. R., Lee, J.-F., Yin, S.-J., Ho, H.-C., Edenberg, H. J., Lu, R.-B., and Kidd, K. K. (1999). Linkage disequilibrium at the ADH1B and ADH1C loci and risk of alcoholism. *Am. J. Hum. Genet.* **64**:1147–1157.
- Osier, M. V., Pakstis, A. J., Soodyall, H., Comas, D., Goldman, D., Odunsi, A., Okonofua, F., Parnas, J., Schulz, L., Bertranpetti, O., Bonne-Tamir, B., Lu, R. B., Kidd, J. R., and Kidd, K. K. (2002). A global perspective on genetic variation at the ADH genes reveals unusual patterns of linkage disequilibrium and diversity. *Am. J. Hum. Genet.* **71**:84–99.
- Osier, M. V., Lu, R.-B., Pakstis, A. J., Kidd, J. R., Huang, S.-Y., and Kidd, K. K. (2004). Possible epistatic role of ADH7 in the protection against alcoholism. *Am. J. Med. Genet. (Neropsychiatric Genetics)* **126**:19–22.
- Reich, T., Edenberg, H. J., Goate, A., Williams, J. T., Rice, J. P., Erdewegh, P. Van, Foroud, T., Hesselbrock, V., Schuckit, M. A., Bucholz, K., Porjesz, B., Li, T.-K., Connelly, P. M., Nurnberger, J. I. Jr., Tischfield, J. A., Crowe, R., Cloninger, C. R., Wu, W., Shears, S., Carr, K., Crose, C., Willig, C., and Begleiter, H. (1998). A genome-wide search for genes affecting the risk for alcohol dependence. *Am. J. Med. Genet.* **81**:207–215.
- Risch, N., and Zhang, H. (1995). Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science* **268**:1584–1589.
- Schmitz, S., Cherny, S. S., and Fulker, D. W. (1998). Increase in power through multivariate analyses. *Behav. Genet.* **28**:357–363.
- Schork, N. J., and Greenwood, T. A. (2004). Inherent bias toward the null hypothesis in conventional multipoint non parametric linkage analysis. *Am. J. Hum. Genet.* **74**:306–316.
- Shen, Y. C., Fan, J. H., Edenberg, H. J., Li, T. K., Cui, Y. H., Wang, Y. F., and Tian, C. H. (1997). Polymorphism of ADH and ALDH genes among four ethnic groups in China and effects upon the risk for alcoholism. *Alcohol Clin. Exp. Res.* **21**:1272–1277.
- Smith, M., Hopkinson, D. A., and Harris, H. (1971). Developmental changes and polymorphism in human alcohol dehydrogenase. *Ann. Hum. Genet.* **34**:251–272.
- Smith, M., Hopkinson, D. A., and Harris, H. (1973). Studies on the subunit structure and molecular size of the human dehydrogenase isozymes determined by the different loci, ADH(1), ADH(2), and ADH(3). *Ann. Hum. Genet.* **36**:401–414.
- Svensson, S., Some, M., Lundsjö, A., Helander, A., Cronholm, T., and Höög, J.-O. (1999). Activities of human alcohol dehydrogenases in the metabolic pathways of ethanol and serotonin. *Eur. J. Biochem.* **262**:324–329.
- Terwilliger, J., and Ott, J. (1994). *Handbook of Human Genetic Linkage*. Baltimore: Johns Hopkins University Press.
- Thomasson, H. R., Edenberg, H. J., Crabb, D. W., Mai, X. L., Jerome, X. L., Li, T. K., Wang, S. P., Lin, Y. T., Lu, R. B., and Yin, S. J. (1991). Alcohol and aldehyde dehydrogenase genotypes and alcoholism in Chinese men. *Am. J. Hum. Genet.* **48**:677–681.
- Tsukahara, M., and Yoshida, A. (1989). Chromosomal assignment of the alcohol dehydrogenase cluster locus to human chromosome 4q21–23 by *in situ* hybridization. *Genomics* **46**:218–220.
- Whitfield, J. B., and Martin, N. G. (1994). Alcohol consumption and alcohol pharmacokinetics: interactions within the normal population. *Alcohol Clin. Exp. Res.* **18**:238–243.
- Whitfield, J. B. (1997). Meta-analysis of the effects of ADH genotype on alcohol dependence and alcoholic liver disease. *Alcohol Clin. Exp. Res.* **32**:613–619.
- Whitfield, J. B., Nightingale, B. N., Bucholz, K. K., Madden, P. A. F., Heath, A. C., and Martin, N. G. (1998). ADH genotype affects alcohol use and dependence in European men. *Alcohol Clin. Exp. Res.* **22**:1463–1469.
- Whitfield, J. B., Zhu, G., Duffy, D. L., Birley, A. J., Madden, P. A., Heath, A. C., and Martin, N. G. (2001). Variation in alcohol pharmacokinetics as a risk factor for alcohol dependence. *Alcohol Clin. Exp. Res.* **25**:1257–1263.
- Williams, J. T., Begleiter, H., Porjesz, B., Edenberg, H. J., Foroud, T., Reich, T., Goate, A., Erdewegh, P. Van, Almasy, L., and Blangero, J. (1999). Joint multipoint linkage analysis of multivariate qualitative and quantitative traits. II. alcoholism and event-related potentials. *Am. J. Hum. Genet.* **65**:1148–1160.
- Yasunami, M., Kikuchi, I., Sarapata, D., and Yoshida, A. (1990). The human Class I alcohol dehydrogenase gene cluster: three genes are tandemly organised in an 89-kb-long segment of the genome. *Genomics* **7**:152–158.
- Zhi, X., Chan, E. M., and Edenberg, H. J. (2000). Tissue-specific regulatory elements in the human alcohol dehydrogenase 6 gene. *DNA Cell. Biol.* **19**:498–497.
- Zgombic-Knight, M., Foglio, M. H., and Duester, G. (1995). Genomic structure and expression of the ADH7 gene encoding human class IV alcohol dehydrogenase, the form most efficient for retinol metabolism in vitro. *J. Biol. Chem.* **270**:4305–4311.

Edited by Richard Rose