# Genome-wide association study of alcohol dependence

**Jens Treutlein, PhD**[1,21,*], **Sven Cichon, PhD**[2,3,*], **Monika Ridinger, MD**[4,*], **Norbert Wodarz, MD**[4], **Michael Soyka, MD**[5,6], **Peter Zill, PhD**[6], **Wolfgang Maier, MD**[7], **Rainald Moessner, MD**[7], **Wolfgang Gaebel, MD**[8], **Norbert Dahmen, MD**[9], **Christoph Fehr, MD**[9], **Norbert Scherbaum, MD**[10], **Michael Steffens, MD**[11], **Kerstin U. Ludwig, MSc**[2], **Josef Frank, MA**[1], **H.- Erich Wichmann, MD, PhD**[12,13], **Stefan Schreiber, MD**[14,15], **Nico Dragano, PhD**[16], **Wolfgang Sommer, MD, PhD**[17,18], **Fernando Leonardi-Essmann, MA**[17], **Anbarasu Lourdusamy, PhD**[19], **Peter Gebicke-Haerter, PhD**[17], **Thomas F. Wienker, MD**[11], **Patrick F. Sullivan, MD**[20], **Markus M. Nöthen, MD**[2,3], **Falk Kiefer, MD**[21], **Rainer Spanagel, PhD**[17,*], **Karl Mann, MD**[21,*], and **Marcella Rietschel, MD**[1,*]

[1]Dept. of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health

[2]Dept. of Genomics, Life & Brain Center, Univ. of Bonn

[3]Inst. of Human Genetics, Univ. of Bonn

[4]Dept. of Psychiatry, Mental State Hospital, Univ. of Regensburg

[5]Private Hospital Meiringen

[6]Dept. of Psychiatry, Univ. of München

[7]Dept. of Psychiatry, Univ. of Bonn

[8]Dept. of Psychiatry and Psychotherapy, Univ. of Düsseldorf

[9]Dept. of Psychiatry, Univ. of Mainz

[10]Addiction Research Group at the Department of Psychiatry and Psychotherapy, Rhine State Hospital Essen, University of Duisburg-Essen

[11]Inst. for Medical Biometry, Informatics and Epidemiology, Univ. of Bonn

[12]Inst. of Epidemiology, Helmholtz Zentrum München

[13]IBE, Chair of Epidemiology, Ludwig Maximilians University Munich

[14]Inst. for Clinical Molecular Biology, Univ. of Kiel

[15]Dept. of General Internal Medicine, Univ. of Kiel

[16]Department of Medical Sociology, University of Düsseldorf for the Heinz Nixdorf Study Group

[17]Dept. of Psychopharmacology, Central Institute of Mental Health

[18]NIAAA, Bethesda

[19]Dipartimento di Medicina Sperimentale e Sanita' Pubblica, Università di Camerino

[20]Department of Genetics, UNC Chapel Hill

[21]Dept. of Addictive Behavior and Addiction Medicine, Central Institute of Mental Health

## Abstract

**Context**—Identification of genes contributing to alcohol dependence will improve our understanding of the mechanisms underlying this disorder.

**Objective**—To identify susceptibility genes for alcohol dependence through a genome-wide association study (GWAS) and follow-up study in a population of German male inpatients with an early age at onset.

**Design**—The GWAS included 487 male inpatients with DSM-IV alcohol dependence with an age at onset below 28 years and 1,358 population based control individuals. The follow-up study included 1,024 male inpatients and 996 age-matched male controls. All subjects were of German descent. The GWAS tested 524,396 single nucleotide polymorphisms (SNPs). All SNPs with $p<10^{-4}$ were subjected to the follow-up study. In addition, nominally significant SNPs from those genes that had also shown expression changes in rat brains after chronic alcohol consumption were selected for the follow-up step.

**Results**—The GWAS produced 121 SNPs with nominal $p<10^{-4}$. These, together with 19 additional SNPs from homologs of rat genes showing differential expression, were genotyped in the follow-up sample. Fifteen SNPs showed significant association with the same allele as in the GWAS. In the combined analysis, two closely linked intergenic SNPs met genome-wide significance (rs7590720 $p=9.72\times10^{-9}$; rs1344694 $p=1.69\times10^{-8}$). They are located on chromosome 2q35, a region which has been implicated in linkage studies for alcohol phenotypes. Nine SNPs were located in genes, including *CDH13* and *ADH1C* genes which have been reported to be associated with alcohol dependence.

**Conclusion**—This is the first GWAS and follow-up study to identify a genome-wide significant association in alcohol dependence. Further independent studies are required to confirm these findings.

## Introduction

Alcohol dependence is characterised by a cluster of cognitive, behavioural and physiological symptoms, with an affected individual continuing to drink despite significant alcohol-induced impairment or distress. Since alcohol affects most human organs, its abuse is associated with a wide range of physical, mental and social harm. According to the World Health Organisation,[1] alcohol abuse constitutes a serious public health problem worldwide, accounting for 4% of the global burden, a burden comparable to the death and disability attributable to tobacco and hypertension.[1,2]

Alcohol dependence is a phenotypically heterogeneous disorder which runs in families and which has a high genetic loading. Twin and adoption studies have shown that 40-60% of the inter-individual phenotypic variance is accounted for by genetic factors.[3-5] In congruence with the observed phenotypic heterogeneity of alcohol dependence, vulnerability to alcohol dependence on the molecular level is thought to be mediated by many genetic loci of small to modest effects in Europeans.[6-14] Despite strenuous efforts and numerous linkage and candidate gene studies, identification of the underlying susceptibility genes has proven difficult. Genome-wide association studies (GWAS) conducted in other complex disorders have been shown to be a successful tool in identifying underlying susceptibility genes (for all published GWAS see: http://www.genome.gov/26525384) and have already resulted in the identification of genetic susceptibility variants for psychiatric disorders such as nicotine addiction, schizophrenia and bipolar disorder[15-24] which were mostly not detected through other diagnosis based discovery approaches. For alcohol dependence, only one GWAS using pooled DNA samples has been published to date, and it has proposed several new susceptibility loci for alcohol dependence. Their products are implicated in cellular signalling, gene regulation, development and cell adhesion.[25] Convergent translational approaches integrate genetic findings from animal models with a candidate gene or GWAS approach in humans and have proven to be very successful.[26,27] The goal of the present study was to conduct a GWAS and follow-up study for alcohol dependence using individual genotyping in samples stratified for homogeneity with respect to sex, ethnicity, age at onset, and recruitment procedures. In order to increase the explanatory power of our findings, we applied a Convergent Functional Genomics[28] approach to integrate findings from gene expression data in alcohol dependent rats with our GWAS findings.

## Methods

### Human Studies

**Patients**—Patients were recruited from consecutive admissions to the psychiatric and addiction medicine departments of 5 different study centers at University hospitals across Southern and Central Germany: Regensburg, Mannheim, Munich, Bonn/Essen/Düsseldorf/Homburg and Mainz. These centres are members of the German Addiction Research Network (GARN; http://www.bw-suchtweb.de) for which one of the co-authors (KM) is spokesman. All patients included in the study suffered from alcohol dependence of such severity that hospitalisation for the treatment or prevention of severe withdrawal symptoms was warranted. All patients received a diagnosis of alcohol dependence (DSM-IV) by consensus of two clinical psychiatrists. Within the GARN study, DSM-IV criteria were

ascertained systematically through use of semi-structured and independently rated interviews conducted by trained staff members.

In Munich and Bonn, patients were assessed using the Semi Structured Assessment for Genetics of Alcoholism (SSAGA),[29] in Mainz the Composite International Diagnostic Interview (CIDI)[30] was used, in Regensburg, Essen/Düsseldorf/Homburg and Mannheim the Structured Clinical Interview for DSM[31] was used. The latter tool was also applied in Munich. All patients and controls were of self reported German ancestry. The study was approved by all relevant local Ethics Committees and all subjects provided written informed consent. In order to increase the power of the study, we increased the homogeneity of the patient sample by including male patients only. Furthermore, only subjects with an early age at onset of alcohol dependence were included, a feature which has a higher heritability in males.[32,33]Patients included in the GWAS (Mannheim N=98; Bonn/Essen/Düsseldorf/Homburg N=30; Mainz N=30; Munich N=53; and Regensburg N=276) had an age at onset <28 (median=20, mean±std=21±4.1) years, defined as the age at which DSM-IV criteria for alcohol dependence were fulfilled for the first time.

The follow-up sample consisted of 1,024 German males recruited at the same sites and under the same protocol as the patients who were entered in the GWAS (Mannheim N=256; Bonn/Essen/Düsseldorf/Homburg N=149; Mainz N=63; Munich N=136; Regensburg N=420). Since most of the patients with an earlier age at onset had already been included in the GWAS, age at onset for the follow-up study was necessarily less stringent and defined as < 45 (median=30, mean±std=29.3±7) years. Phenotypes and genotypes resulting from GWAS are stored in a comprehensive database at the Institute for Medical Biometry, Informatics and Epidemiology in Bonn (IMBIE).

**Control subjects**—A total of 1,358 control subjects were included in the GWAS. Controls were taken from three population-based epidemiological studies: 1) N=487 from PopGen[34] N=488 from KORA[35] and 3) N=383 from Heinz Nixdorf Recall Study (HNR)[36]. These three recruitment areas are located in Schleswig-Holstein (North Germany), Augsburg (South Germany) and Essen, Bochum, Mülheim (Ruhr-Area, West Germany). The PopGen (Population-based Recruitment of Patients and Controls for the Analysis) project (http://www.science.ngfn.de/10_233.htm; www.popgen.de;) was initiated to provide all locally prevalent cases with the diseases in question for the disease-orientated projects from the NGFN as well as population-based control samples which will ultimately comprise 7200 persons. The KORA project (Cooperative Health Research in the Region of Augsburg http://www.science.ngfn.de/10_234.htm; www.gsf.de/KORA), which has evolved from the WHO MONICA study (Monitoring of Trends and Determinants of Cardiovascular Disease), has the biosamples, phenotypic characteristics and environmental parameters of 18,000 adults from Augsburg and the surrounding counties. The biological specimen bank was established in order to enable researchers to perform epidemiological research into molecular and genetic factors. The Heinz Nixdorf Recall (Risk Factors, Evaluation of Coronary Calcification, and Lifestyle) study (http://www.recall-studie.uni-essen.de) was started in 2000 as a prospective cohort study to determine predictors of coronary heart disease. It has the biosamples, phenotypic characteristics and environmental parameters of 4500 adults. PopGen and Kora are the two major German biobanking resources available to participants

of the National Genome Research Project (NGFN) (http://www.ngfn.de/) for use as universal controls,[37] and they have already been used as samples in other GWAS.[37-38]

Male controls for the follow-up study were drawn from the KORA study (N=382) and from a population based sample (N=614) which had been collected in the area of Bonn by one of the authors (MR) for use as a control sample for association studies within the framework of the NGFN. Controls were not stratified for alcohol abuse/dependence, which may introduce a conservative bias. Tests for genetic differentiation between German populations have shown low levels of population substructure, thus demonstrating that the German population is an appropriate source for use in association studies of complex diseases.[40]

**DNA preparation, genotyping and statistical analysis**—Genomic DNA was prepared from whole blood according to standard procedures. For all samples, DNA concentrations were quantified in triplicate measurements by Picogreends DNA Quantitation Kit (www.invitrogen.com) and normalized to 50 ng/ul at the CIMH, Molecular Genetics Laboratory. For the GWAS, all samples were genotyped individually using Human Hap 550 BeadChips (Illumina, Inc., San Diego, CA, USA). The patient and the PopGen and KORA samples were genotyped at Illumina Inc. and the HNR sample was genotyped at the Department of Genomics at the Life & BrainCenter, University of Bonn. Genotyping of the follow-up sample was performed by primer extension reaction chemistry with MALDI-TOF mass spectrometry using the iPLEX® Assay (Sequenom, San Diego, USA) at the Life & BrainCenter, University of Bonn. Genotyping results were imported into a central computer system for statistical analysis.

**Data analysis and quality control**—Data analysis and quality control (QC) was performed using the software packages R version 2.5.1 (http://www.R-project.org) and PLINK version 1.0.3.[41] In the GWAS, genotype data were cleaned before analysis by removing SNPs or individuals that did not fulfil the QC criteria, which included: SNP call proportion 95%, subject completeness proportion 95%, SNP minor allele frequency 0.01 and SNP conformity with Hardy-Weinberg equilibrium expectations (p 0.01 in controls). To correct for cryptic relatedness, all pairs of individuals displaying an identity-by-state (IBS) value larger than 1.6 were marked, and for each pair the individual with the lower typing rate was removed from the analysis. Cochran Armitage trend statistics were used to calculate significant association for autosomal SNPs. To visualize the outcome of the QC steps, Cochran Armitage p-values were depicted in a quantile-quantile plot (Figure 1). We observed good adherence of p-values to the line of expectance, which implies that potential spurious associations characterized by an inflation of highly significant p-values were successfully removed by our QC measures. The remaining slight deviations from the line of expectance are interpreted to include true genetic effects. Further correction for lambda[42] improved the QQ-plot of Armitage p-values (eFigure1).

### Animal Studies

**Animals**—Three groups of 2-3 months old alcohol preferring rats were used for long-term alcohol consumption and gene expression profiling: male P rats (n = 15; IndianaUniversity, Indianapolis), male HAD rats (n = 13; IndianaUniversity, Indianapolis) and male AA rats

(n=14; National Public Health Institute, Helsinki). The rats were kindly provided by T.K. Li (Department of Psychiatry, Institute of Psychiatric Research, Indiana University School of Medicine, Indianapolis)and D. Sinclair (Department of Mental Health and Alcohol Research, National Public Health Institute, Helsinki). Each rat strain shows alcohol preference due to various neurochemical alterations, as indicated by the abbreviations P (preference), HAD (high alcohol drinking), AA (alcohol addicted).[43,44] All experimental procedures were approved by the Committee on Animal Care and Use (Regierungspräsidium Karlsruhe), and carried out in accordance with the local Animal Welfare Act and the European Communities Council Directive of 24 November 1986 (86/609/EEC).

According to the protocol of Vengeliene et al.,[45,46] 8 P rats, 7 HAD rats, and 7 AA rats were given *ad libitum* access to tap water and to 5%, and 20% ethanol solution (v/v). All rats underwent a two-week deprivation cycle after 8 weeks of continuous alcohol availability. After the deprivation period, rats were given access to alcohol again and 3 more two-week deprivation periods were introduced in a random manner (the duration between deprivation periods varied between 4 and 16 weeks). The long-term voluntary alcohol drinking procedure, including all deprivation phases, lasted a total of 52 weeks. Total ethanol intake (g/kg of body weight/day) was calculated as the daily average across 7 measuring days. For comparison, 3 age- and weight-matched control groups, consisting of 7 P rats, 6 HAD rats, and 7 AA rats, experienced identical handling procedures for the entire duration of the experiment, but did not receive alcohol.

**Gene expression profiling in alcohol-preferring rat strains—**Preparation of brain samples and RNA isolation are described in the Supplemental Text. Target preparation was done for individual samples from caudate putamen and amygdala using 5 μg of total RNA. Hybridization to RG U34A arrays, staining, washing and scanning of the chips were performed according to the manufacturer's technical manual (Affymetrix, Santa Clara, CA).

**Data mining—**Micro Array Suite 5.0 (Affymetrix) derived cell intensity files (CEL) were processed in R 2.1.1 language and environment (http://www.R-project.org) using Bioconductor 1.6 packages.[47] Each array was inspected for regional hybridization bias and quality control parameters as recently described.[48] Fifty-three arrays (27 from caudate putamen and 26 from amygdala) passed the quality filter and were included in the statistical analysis. Of the 8799 probe sets on the RG_U34A array, only those with intensity values > 100 in at least 25 % of the samples were retained (6344 probe sets). A 3-way ANOVA was used to identify differentially expressed genes across strain, brain region and treatment. The list of genes affected by long-term ethanol consumption included those with a $p < 0.05$ for treatment from the 3-way ANOVA. Posthoc analysis was performed via template matching across strains. Correlation coefficients (r) for consistent up or down regulation by ethanol in either caudate putamen or amygdala or both regions were calculated.

**Integration of animal data into the human study—**SNPs which have been found to be associated with the disorder in the GWAS with $p<1\times10-3$ or which were found to lie in a SNP cluster (cluster definition: 2 SNPs with p 1×10-2 and at least one SNP with $p<1\times10-3$ in a distance of <30kb) (see, e.g., eFigures 2-4) were taken forward to the follow-

up study if they were located in a homolog of a rat gene showing differential brain expression with long-term high ethanol consumption compared to their respective controls without ethanol access. Only one SNP was selected for each gene. If more than one SNP in the gene fulfilled the above mentioned criteria, the SNP with the highest number of assigned transcripts (Sullivan et al., annotation file: https://slep.unc.edu/evidence/) was taken forward. In cases of similar transcript numbers, the SNP with the lowest p-value was chosen.

In the animal study, we have used a convergent approach with two strategic lines to integrate the rat data:

  **a.** a qualitative approach looking for orthologous or paralogous genes in GWAS and expression profiling in rats (see list of genes in eTables 2 and 3).

  **b.** a (semi-quantitative) ranking strategy as theoretically described by Bertsch et al.,[28] where those orthologous or paralogous genes were included. In this strategy a weighing ratio of 1 (both datasets with equal weights) was employed, multiplying the p-values of respective rat and human genes.

With strategy a) a total number of 19 additional SNPs were identified and ranked according to strategy b). Since the total number of SNPs derived through the animal approach was only 19, we eventually decided to include all of them in the follow-up study instead of selecting a small subset based on the ranking order of genes. The three genes which were confirmed by the follow up study (Tab. 1) had been ranked # 7 (ADH), #12 (CDH), and # 19 (GATA) in approach b). We want to conclude from these data, that it is hard to introduce quantitative evaluations in those convergent approaches. Nevertheless, convergent approaches as pursued in this study can provide very valuable additional information and apparently become increasingly more popular.

We searched Gene Ontology (http://amigo.geneontology.org/cgi-bin/amigo/go.cgi) for the annotation of gene products for those genes in which confirmed SNPs of the follow up study are located. Gene Ontology provides a standardised vocabulary describing gene products in order to ensure uniformity across databases.[49] We investigated whether these genes share distribution patterns in Gene Ontology (GO) categories by mapping them to GO annotations.

## Results

### Human Studies

We excluded 11 of the total 487 GWAS cases with alcohol dependence from further analyses. Of these, six DNA samples failed to genotype, and, in five subjects, the genome-wide IBS score was above 1.60, indicating possible cryptic relatedness. Of the 561,466 SNPs genotyped per individual, 99.6% passed quality control (QC). The final data set was comprised of 524,396 SNPs (511,701 autosomal and 12,695 sex-linked markers) in a total of 476 male cases and 1,358 male controls. The mean call rate of the final set of SNPs (cases and controls) was 99.76% (±0.45%).

We used two complementary SNP selection strategies, a 'lowest p-value' strategy and a 'rodent candidate gene' strategy, to prioritize SNPs for follow-up. In total, 139 SNPs were carried forward for genotyping in the follow-up sample. These included all SNPs with a p-

value $<10^{-4}$ (Armitage trend test for autosomes and allelic test for x-chromosome), n=121, of which an assay design was possible for n=120 (see Supplementary Online Content eTable 1), and additional 19 SNPs with at least nominal significance, which would not have been accounted for by the 'lowest p-value' approach. These 19 SNPs are among 22 SNPs (three of which were already represented in the aforementioned 'lowest p-value' selection) located in human homologs of rat genes showing differential expression in the rat brain following chronic alcohol consumption (for more details see Supplementary Online Content) and therefore have a higher *a priori* probability of being involved in the aetiology of alcohol dependence.

In the follow-up study, 16 SNPs showed association with at least nominal significance (p<0.05, two-sided), 15 of which (9 intragenic, 6 intergenic) were associated with the same allele as in the GWA study. The number of significantly replicated SNPs is higher than expected by chance (p=$2.45\times10^{-6}$). Three of the intragenic SNPs are derived from the 22 SNPs selected on the basis of the animal model (Table 1). This number is also significantly higher than that expected by chance (p=$1.69\times10^{-2}$).

Combining the data across the two samples showed genome-wide significance (Bonferroni adjusted α-level of 0.05/524k)[50,51] for two SNPs, i.e. rs7590720 (p=$9.72\times10^{-9}$) and rs1344694 (p=$1.69\times10^{-8}$). These SNPs map to the 3′-flanking region of the gene encoding peroxisomal trans-2-enoyl-coA reductase (*PECR*), located on chromosome 2q35. They are in linkage disequilibrium with each other and with another SNP, rs705648 (p=$1.78\times10^{-6}$),located within the *PECR* gene (rs1344694 – rs7590720: D′=1.0, $r^2$=0.739; rs1344694 - rs705648: D'= 0.943, $r^2$=0.568; rs7590720 – rs705648: D′=0.948, $r^2$=0.776; HapMap).

A further eight SNPs are located within genes: rs13362120 (combined p-value $1.85\times10^{-5}$) in the calpastatin (*CAST*) gene; rs13160562 (combined p-value $7.09\times10^{-6}$) in the endoplasmatic reticulum aminopeptidase 1 (*ERAP1*)/calpastatin (*CAST*) genes; rs1864982 (combined p-value $3.46\times10^{-6}$) in the protein phosphatase 2 (formerly 2A), regulatory subunit B, beta isoform (*PPP2R2B*)gene; rs6902771 (combined p-value $8.30\times10^{-6}$) in the estrogen receptor 1 (*ESR1*) gene; rs7138291 (combined p-value $3.68\times10^{-5}$) in the coiled-coil domain containing 41 (*CCDC41*) gene; rs1614972 (combined p-value $1.41\times10^{-4}$) in the alcohol dehydrogenase 1c (*ADH1C*) gene; rs13273672 (combined p-value $4.75\times10^{-4}$) located in the GATA binding protein 4 (*GATA4*) gene and rs11640875 (combined p-value $1.84\times10^{-5}$) in the cadherin 13 (*CDH13*) gene. The first five SNPs were selected for the follow-up study through the 'lowest p-value' strategy and the latter three via the 'rodent candidate gene' approach. These findings and those of the intergenic SNPs are provided in Table 1, Table 2 and eTable 5.

Specific GO biological process, cellular component and molecular function terms remained largely unique for each of the genes and were not found to be enriched among the genes analysed. No specific coherent underlying process could be deduced from GO term associations. As a consequence, we considered the literature-derived molecular function of each of the gene products of intragenic SNPs separately.

### Animal Studies

The GeneChip® Rat Genome U34 Set provides gene expression data for more than 24,000 known genes and EST clusters for comprehensive coverage of the rat genome. The two brain regions were evaluated separately. 3-way ANOVA revealed 542 genes differentially expressed at a $p<0.05$ in either caudate putamen or amygdala or both brain regions of alcohol preferring P, HAD, and AA rats after one year of ethanol consumption. A detailed description of the 'rodent candidate gene' strategy for selection of markers for the follow-up study is given in paragraph 'Integration of animal data into the human study'.

## Comment

Our GWA and replication study of alcohol dependence has detected and replicated evidence for association with 15 markers. Two of these markers, rs7590720 and rs1344694, remained significant after genome-wide correction for multiple testing in the combined sample of 1460 patients and 2332 control persons.

These two markers are located ∼5kb apart in chromosomal region 2q35, which has been implicated in alcohol dependence through previous linkage studies. Linkage to this region was found in a genome-wide search in 2282 individuals from 262 families with a high density of alcohol dependence by the Collaborative Study on the Genetics of Alcoholism (COGA).[52] Near the microsatellite D2S1371, the highest LOD score for comorbid alcoholism and depression phenotype was 4.12 (this LOD score however, was only obtained in one dataset, i.e. in the replication dataset, while in the initial and combined datasets the LOD scores were 0.00 and 2.16, respectively).[52] Marker D2S1371 is located ∼1.4Mb away from rs7590720 and rs1344694. Another COGA project[53] found linkage with a maximum LOD score of 2.4 between a low level of response (LR) to alcohol and the microsatellite marker D2S434 which is only ∼1.7Mb away from our markers rs7590720 and rs1344694. The LR to alcohol is an endophenotype related to heavy drinking and alcohol problems.[54] It is genetically influenced in both animals[55-57] and humans[58,59] with an estimated heritability in humans of between 40-50%[60]. Two further linkage studies of the COGA project[61,62] reported linkage with LOD scores of 3.28 and 3.0, respectively, between the same marker D2S434 and the P3(00) component of the event related potential (ERP). Heritability of the P3 amplitude ranges between 50-80% (for review see [62]) and P3 deficits have been repeatedly reported to be a strong vulnerability factor for alcohol dependence, especially in males with a high familial loading of alcohol dependence.[63,64] In a linkage scan conducted by Hill et al.[65] in 330 members from 65 families with a high density of alcohol dependence, linkage to 2q35 with marker D2S2382 was identified. This marker is located only ∼0.2Mb from our two main SNPs. A LOD score of 3.68 was observed in a four-parameter model which included measurement of age, gender and alcohol dependence along with measurement of constraint - a personality construct[66] which appears to overlap with risk taking behaviour.

These linkage findings, which all satisfied the Lander and Kruglyak[67] criteria for suggestive linkage, underline the potential importance of this chromosomal region. However, since no meta-analysis of linkage studies in alcohol addiction exists, the possibility that these findings are merely due to random noise cannot be ruled out.

The gene which is closest to the peak association signal and which also harbours another associated SNP, rs705648 (p=$1.78\times10^{-6}$), is the peroxisomal trans-2-enoyl-coA reductase (*PECR*). *PECR* appears to be a key enzyme of fatty acid metabolism as it catalyzes the reduction of medium chain (unsaturated) enoyl-CoAs to saturated acyl-CoAs which may then undergo oxidation in mitochondria for energy production. This pathway, originating from triglycerides, is particularly important in conditions of starvation when energy supply is switched from utilization of glucose to fatty acids. In this condition, PECR needs to be up-regulated. Triglycerides are typically increased in the blood of alcohol dependent individuals.[68,69] PECR reduces C-C double bonds in not only even-chain, but also in odd-chain fatty acids. On degradation to acetyl-CoA, odd-chain fatty acids end up with propionyl-CoA, which requires carboxylation to an even-chain fatty acid. This step is mediated through methyl-malonyl-CoA, an intermediate formed by the essential cofactors methyl-tetrahydrofolate as the methyl-donor, and vitamin B12. Folate deficiency, often observed in alcohol-dependent patients,[70] may therefore interfere with this pathway and prevent the availability of a sufficient supply of acetyl-CoA. Expression of *PECR* is highest in the liver, followed by the kidney, muscle tissue, the lungs and the heart, but barely – if at all – in the brain.[71] If this gene is involved in alcohol dependence, it must act in the periphery rather than in the CNS. It could therefore be expected that we detected *PECR* only in our GWAS, which identifies susceptibility genes independent of organ-specific expression, but not in the animal approach, which is brain-specific.

A further 4 SNPs are intergenic and further eight SNPs are located in genes. No ontological relatedness was detected between these genes. That is not surprising, given the assumed genetic heterogeneity of alcohol addiction. In the Supplemental Text, we discuss data from linkage findings, biochemistry and animal studies that have provided evidence that these loci may be involved in alcohol dependence. However, none of these markers achieved genome-wide significance in our study. This does not necessarily disqualify them. Alcohol dependence is a complex and heterogeneous disorder that is influenced by multiple genetic, biological, psychological, and sociocultural factors. Formal genetic and linkage studies do not indicate the existence of genes with major effects in the European population. Thus, genome-wide significant findings in moderately-sized samples will be rare, but the number of replicable associations may actually be much larger. One replicated association that did not withstand correction for multiple testing was obtained with the SNP rs1614972 in the *ADH1C* gene. The *ADH* gene cluster is located on chromosomal region 4q, which is among the most consistently replicated loci contributing to alcohol phenotypes.[72] The *ADH1C* gene belongs to the biologically relevant alcohol-metabolising alcohol dehydrogenase (*ADH*) genes. The p-value of this SNP, which was at position 282 of our top-down ranking, achieved a p=$2.84\times10^{-4}$ in the GWAS. It would not have been carried forward into the follow-up study if it had not been supported by evidence from our animal study. In order to achieve a power of 80% to detect significant association withstanding correction for multiple testing, we would have needed to include 5573 patients and 8900 controls (regarding actual OR and MAF of this SNP). The same is true for our replicated association to marker rs11640875 which had rank no. 510 in our GWAS findings. It is located in the cadherin 13 (*CDH13*) gene, a gene which has been described in previous studies as a susceptibility locus for both alcohol dependence[25] and methamphetamine dependence.[73]

Increasing sample sizes is often not feasible and bears the risk of increasing clinical as well as genetic heterogeneity. Experience from GWAS into diabetes, for example, has shown that differences in assessment can jeopardize even the most robust findings.[74] Obtaining very large sample sizes may necessitate inclusion of samples from different cultural and ethnic backgrounds which can also increase heterogeneity. The *ADH1C* gene belongs to the *ADH*class I genes (*ADH1A*, *ADH1B* and *ADH1C*) which have been extensively investigated for the risk of alcohol dependence, initially in Asian[75,76,77]and subsequently in African[78,79]and European[80-84]populations. A metaanalysis[85] reported a significantly higher risk of alcohol dependence among carriers of the ADH1C*2 allele (OR=1.91) in East Asian populations, but its role in the European population is less clear. Our associated marker rs1614972 is in complete LD (D'=1.0; $r^2$=0.311, HapMap) with one of the two polymorphisms of the ADH1C*2 protein isoform that has recently been investigated in 575 patients and 530 controls from the Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD) sample. The two marker haplotype comprised of markers rs1614972 and rs1693482 yielded evidence for association with alcohol dependence. Single marker analysis with rs1614972, however, produced no evidence at all. While the allele frequencies in Irish and German patients are the same, they differ between the Irish and German control samples; the allele frequencies between the German GWAS and the follow-up study being the same. This finding underlines the importance of sample homogeneity.

In order to increase the genetic homogeneity of the sample, and thus increase the power of the study, we attempted to select patients as rigorously as possible. Patients of German descent were recruited within a joint research project and were stratified for male gender and an early age at onset, a combination of traits which is associated with increased heritability.[32,33] Prior studies have shown that population stratification in Germansare small.[40,86,87] Thus, at the time when SNPs for the follow-up study were chosen, we had not applied a formal test to control for population stratification. A post-hoc principal components analysis (PCA)[88](EigenSoftv1.01; http://www.genepath.med.harvard.edu/~reich/Software.htm) and genomic control analysis[42] (showing a lambda of 1.099) had no major impact on the significance level of our 15 SNPs (see Supplementary Online Content, eTable 4), thus confirming these prior obervations.

Alcohol dependence as defined by DSM-IV is complex, and association studies with more biologically refined (endo) phenotypes will help to disentangle this heterogeneity. Our approach of selecting more genetically homogeneous groups by stratifying for male sex and and an earlier age at onset is a first step in this direction. We must point out, however, that in order to achieve our overall sample of more than 1500 patients, it was necessary to relax our early age at onset criteria. While in the GWAS patients the upper inclusion limit was 28 years with a median of 20 years, the upper inclusion limit in the replication sample had to be extended to 45 years with a median of 30 years. Although, this cannot be considered as a specifically early age at onset, the sample as a whole and especially the GWAS sample has clearly been stratified for early age at onset.

In conclusion, we have identified genome-wide significant association with two SNPs located in a chromosomal region for which linkage and animal studies provide compelling evidence for the presence of susceptibility variants predisposing to alcohol addiction.

Further independent studies are required to confirm these findings. The gene, which is located close to these SNPs - *PECR*- is a plausible candidate and merits further investigation. We have shown that the GWA approach is a powerful tool for producing valuable findings in even a moderately-sized sample, when it has been carefully selected for genetically homogeneous cases. In addition, our study underlines the value of a convergent approach between animal studies and systematic genetic screens in humans in order to deal with the multiple testing problem in GWAS.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

## References

1. World Health Organisation, World Health Report. Reducing Risks, Promoting Healthy Life. Geneva: WHO; 2002.

2. World Health Organisation. WHO Global Status Report on Alcohol 2004. World Health organisation Department of Mental Health and Substance Abuse; Geneva: 2004.

3. Radel M, Goldman D. Pharmacogenetics of alcohol response and alcoholism: the interplay of genes and environmental factors in thresholds for alcoholism. Drug Metab Dispos. 2001; 29(4 Pt 2):489–494. [PubMed: 11259338]

4. Tyndale RF. Genetics of alcohol and tobacco use in humans. Ann Med. 2003; 35(2):94–121. 2003. [PubMed: 12795339]

5. Goldman D, Oroszi G, Ducci F. The genetics of addictions: uncovering the genes. Nat Rev Genet. 2005; 6(7):521–532. [PubMed: 15995696]

6. Cloninger CR. Neurogenetic adaptive mechanisms in alcoholism. Science. 1987; 24(4800):236. 410–416.

7. Enoch MA. Pharmacogenomics of alcohol response and addiction. Am J Pharmacogenomics. 2003; 3(4):217–232. [PubMed: 12930156]

8. Dick DM, Foroud T. Candidate genes for alcohol dependence: a review of genetic evidence from human studies. Alcohol ClinExp Res. 2003; 27(5):868–879.

9. Dick DM, Bierut LJ. The genetics of alcohol dependence. Curr Psychiatry Rep. 2006; 8(2):151–157. [PubMed: 16539893]

10. Higuchi S, Matsushita S, Kashima H. New findings on the genetic influences on alcohol use and dependence. Curr Opin Psychiatry. 2006; 19(3):253–265. [PubMed: 16612210]

11. Edenberg HJ, Foroud T. The genetics of alcoholism: identifying specific genes through family studies. Addict Biol. 2006; 11(3-4):386–396. [PubMed: 16961766]

12. Porjesz B, Rangaswamy M. Neurophysiological endophenotypes, CNS disinhibition, and risk for alcohol dependence and related disorders. ScientificWorldJournal. 2007; 7:131–41. [PubMed: 17982586]

13. Gorwood, P.; Schumann, G.; Treutlein, J.; Ades, J. Pharmacogenetics of alcohol dependence. In: Gorwood, P.; Hamon, M., editors. Psychopharmacogenetics. Heidelberg, Germany: Springer; p. 2006p. 177-202.

14. Mayfield RD, Harris RA, Schuckit MA. Genetic factors influencing alcohol dependence. Br J Pharmacol. 2008; 154(2):275–287. [PubMed: 18362899]

15. Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, Manolescu A, Thorleifsson G, Stefansson H, Ingason A, Stacey SN, Bergthorsson JT, Thorlacius S, Gudmundsson J, Jonsson T, Jakobsdottir M, Saemundsdottir J, Olafsdottir O, Gudmundsson LJ, Bjornsdottir G, Kristjansson K, Skuladottir H, Isaksson HJ, Gudbjartsson T, Jones GT, Mueller T, Gottsäter A, Flex A, Aben KK, de Vegt F, Mulders PF, Isla D, Vidal MJ, Asin L, Saez B, Murillo L, Blondal T, Kolbeinsson H, Stefansson JG, Hansdottir I, Runarsdottir V, Pola R, Lindblad B, van Rij AM, Dieplinger B, Haltmayer M, Mayordomo JI, Kiemeney LA, Matthiasson SE, Oskarsson H, Tyrfingsson T, Gudbjartsson DF, Gulcher JR, Jonsson S, Thorsteinsdottir U, Kong A, Stefansson K. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature. 2008; 452(7187):638–642. 2008. [PubMed: 18385739]

16. Uhl GR, Liu QR, Drgon T, Johnson C, Walther D, Rose JE, David SP, Niaura R, Lerman C. Molecular genetics of successful smoking cessation: convergent genome-wide association study results. Arch Gen Psychiatry. 2008; 65(6):683–693. [PubMed: 18519826]

17. Mah S, Nelson MR, Delisi LE, Reneland RH, Markward N, James MR, Nyholt DR, Hayward N, Handoko H, Mowry B, Kammerer S, Braun A. Identification of the semaphorin receptor *PLXNA2* as a candidate for susceptibility to schizophrenia. Mol Psychiatry. 2006; 11(5):471–478. [PubMed: 16402134]

18. Lencz T, Morgan TV, Athanasiou M, Dain B, Reed CR, Kane JM, Kucherlapati R, Malhotra AK. Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. Mol Psychiatry. 2007; 12(6):572–580. [PubMed: 17522711]

19. Sullivan PF, Lin D, Tzeng JY, van den Oord E, Perkins D, Stroup TS, Wagner M, Lee S, Wright FA, Zou F, Liu W, Downing AM, Lieberman J, Close SL. Genomewide association for schizophrenia in the CATIE study: results of stage 1. Mol Psychiatry. 2008; 13(6):570–584. [PubMed: 18347602]

20. O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V, Nikolov I, Hamshere M, Carroll L, Georgieva L, Dwyer S, Holmans P, Marchini JL, Spencer C, Howie B, Leung HT, Hartmann AM, Möller HJ, Morris DW, Shi Y, Feng G, Hoffmann P, Propping P, Vasilescu C, Maier W, Rietschel M, Zammit S, Schumacher J, Quinn EM, Schulze TG, Williams NM, Giegling I, Iwata N, Ikeda M, Darvasi A, Shifman S, He L, Duan J, Sanders AR, Levinson DF, Gejman PV, Molecular Genetics of Schizophrenia Collaboration. Cichon S, Nöthen MM, Gill M, Corvin A, Rujescu D, Kirov G, Owen MJ. Identification of novel schizophrenia loci by genome-wide association and follow-up. Nat Genet. 2008; 40(9):1053–1055. [PubMed: 18677311]

21. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447(7145):661–678. [PubMed: 17554300]

22. Baum AE, Akula N, Cabanero M, Cardona I, Corona W, Klemens B, Schulze TG, Cichon S, Rietschel M, Nöthen MM, Georgi A, Schumacher J, Schwarz M, Abou JamraR, Höfels S, Propping P, Satagopan J, Detera-Wadleigh SD, Hardy J, McMahon FJ. A genome-wide association study implicates diacylglycerol kinase eta (*DGKH*) and several other genes in the etiology of bipolar disorder. Mol Psychiatry. 2008; 13(2):197–207. [PubMed: 17486107]

23. Sklar P, Smoller JW, Fan J, Ferreira MA, Perlis RH, Chambert K, Nimgaonkar VL, McQueen MB, Faraone SV, Kirby A, de Bakker PI, Ogdie MN, Thase ME, Sachs GS, Todd-Brown K, Gabriel SB, Sougnez C, Gates C, Blumenstiel B, Defelice M, Ardlie KG, Franklin J, Muir WJ, McGhee KA, MacIntyre DJ, McLean A, VanBeck M, McQuillin A, Bass NJ, Robinson M, Lawrence J, Anjorin A, Curtis D, Scolnick EM, Daly MJ, Blackwood DH, Gurling HM, Purcell SM. Whole-genome association study of bipolar disorder. Mol Psychiatry. 2008; 13(6):558–569. [PubMed: 18317468]

24. Ferreira MA, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L, Fan J, Kirov G, Perlis RH, Green EK, Smoller JW, Grozeva D, Stone J, Nikolov I, Chambert K, Hamshere ML, Nimgaonkar VL, Moskvina V, Thase ME, Caesar S, Sachs GS, Franklin J, Gordon-Smith K, Ardlie KG, Gabriel SB, Fraser C, Blumenstiel B, Defelice M, Breen G, Gill M, Morris DW, Elkin A, Muir WJ, McGhee KA, Williamson R, Macintyre DJ, Maclean AW, St Clair D, Robinson M, Van Beck M, Pereira AC, Kandaswamy R, McQuillin A, Collier DA, Bass NJ, Young AH, Lawrence J, Nicol Ferrier I, Anjorin A, Farmer A, Curtis D, Scolnick EM, McGuffin P, Daly MJ,

Corvin AP, Holmans PA, Blackwood DH, Wellcome Trust Case Control Consortium. Gurling HM, Owen MJ, Purcell SM, Sklar P, Craddock N. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. Nat Genet. 2008 Aug 17. Epub ahead of print.

25. Johnson C, Drgon T, Liu QR, Walther D, Edenberg H, Rice J, Foroud T, Uhl GR. Pooled association genome scanning for alcohol dependence using 104,268 SNPs: validation and use to identify alcoholism vulnerability loci in unrelated individuals from the collaborative study on the genetics of alcoholism. Am J Med Genet B Neuropsychiatr Genet. 2006; 141B(8):844–853. [PubMed: 16894614]

26. Spanagel R, Pendyala G, Abarca C, Zghoul T, Sanchis-Segura C, Magnone MC, Lascorz J, Depner M, Holzberg D, Soyka M, Schreiber S, Matsuda F, Lathrop M, Schumann G, Albrecht U. The clock gene Per2 influences the glutamatergic system and modulates alcohol consumption. Nat Med. 2005; 11(1):35–42. [PubMed: 15608650]

27. Rodd ZA, Bertsch BA, Strother WN, Le-Niculescu H, Balaraman Y, Hayden E, Jerome RE, Lumeng L, Nurnberger JI Jr, Edenberg HJ, McBride WJ, Niculescu AB. Candidate genes, pathways and mechanisms for alcoholism: an expanded convergent functional genomics approach. Pharmacogenomics J. 2007; 7(4):222–256. [PubMed: 17033615]

28. Bertsch B, Ogden CA, Sidhu K, Le-Niculescu H, Kuczenski R, Niculescu AB. Convergent functional genomics: a Bayesian candidate gene identification approach for complex disorders. Methods. 2005; 37(3):274–279. [PubMed: 16308156]

29. Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI Jr, Reich T, Schmidt I, Schuckit MA. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. J Stud Alcohol. 1994; 55(2):149–158. 1994. [PubMed: 8189735]

30. Wittchen HU. Reliability and validity studies of the WHO--Composite International Diagnostic Interview (CIDI): a critical review. J Psychiatr Res. 1994; 28(1):57–84. 1994. [PubMed: 8064641]

31. Spitzer, RL.; Williams, JBW.; Gibbon, M.; First, MB. Structured Clinican Interview for DSM-III-R Axis II Disorders,(SCID-II). Washington, D.C.: American Psychiatric Press, Inc; 1990.

32. Pickens RW, Svikis DS, McGue M, Lykken DT, Heston LL, Clayton PJ. Heterogeneity in the inheritance of alcoholism. A study of male and female twins. Arch Gen Psychiatry. 1991; 48(1):19–28. [PubMed: 1984758]

33. Liu IC, Blacker DL, Xu R, Fitzmaurice G, Tsuang MT, Lyons MJ. Genetic and environmental contributions to age of onset of alcohol dependence symptoms in male twins. Addiction. 2004; 99(11):1403–1409. [PubMed: 15500593]

34. Krawczak M, Nikolaus S, von Eberstein H, Croucher PJ, ElMokhtari NE, Schreiber S. PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. Community Genet. 2006; 9(1):55–61. [PubMed: 16490960]

35. Wichmann HE, Gieger C, Illig T. MONICA/KORA Study Group. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen. 2005; 67(Suppl 1):S26–30. [PubMed: 16032514]

36. Schmermund A, Möhlenkamp S, Stang A, Grönemeyer D, Seibel R, Hirche H, Mann K, Siffert W, Lauterbach K, Siegrist J, Jöckel KH, Erbel R. Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL Study. Risk Factors, Evaluation of Coronary Calcium and Lifestyle. Am Heart J. 2002; 144(2):212–218. [PubMed: 12177636]

37. Luca D, Ringquist S, Klei L, Lee AB, Gieger C, Wichmann HE, Schreiber S, Krawczak M, Lu Y, Styche A, Devlin B, Roeder K, Trucco M. On the use of general control samples for genome-wide association studies: genetic matching highlights causal variants. Am J Hum Genet. 2008; 82(2):453–463. [PubMed: 18252225]

38. Winkelmann J, Schormair B, Lichtner P, Ripke S, Xiong L, Jalilzadeh S, Fulda S, Pütz B, Eckstein G, Hauk S, Trenkwalder C, Zimprich A, Stiasny-Kolster K, Oertel W, Bachmann CG, Paulus W, Peglau I, Eisensehr I, Montplaisir J, Turecki G, Rouleau G, Gieger C, Illig T, Wichmann HE, Holsboer F, Müller-Myhsok B, Meitinger T. Genome-wide association study of restless legs

syndrome identifies common variants in three genomic regions. Nat Genet. 2007; 39(8):1000–1006. [PubMed: 17637780]

39. Buch S, Schafmayer C, Völzke H, Becker C, Franke A, von Eller-Eberstein H, Kluck C, Bässmann I, Brosch M, Lammert F, Miquel JF, Nervi F, Wittig M, Rosskopf D, Timm B, Höll C, Seeger M, ElSharawy A, Lu T, Egberts J, Fändrich F, Fölsch UR, Krawczak M, Schreiber S, Nürnberg P, Tepel J, Hampe J. A genome-wide association scan identifies the hepatic cholesterol transporter ABCG8 as a susceptibility factor for human gallstone disease. Nat Genet. 2007; 39(8):995–999. [PubMed: 17632509]

40. Steffens M, Lamina C, Illig T, Bettecken T, Vogler R, Entz P, Suk EK, Toliat MR, Klopp N, Caliebe A, König IR, Köhler K, Ludemann J, Diaz Lacava A, Fimmers R, Lichtner P, Ziegler A, Wolf A, Krawczak M, Nürnberg P, Hampe J, Schreiber S, Meitinger T, Wichmann HE, Roeder K, Wienker TF, Baur MP. SNP-based analysis of genetic substructure in the German population. Hum Hered. 2006; 62(1):20–29. [PubMed: 17003564]

41. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81(3):559–575. [PubMed: 17701901]

42. Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999; 55(4):997–1004. [PubMed: 11315092]

43. Bell RL, Rodd ZA, Lumeng L, Murphy JM, McBride WJ. The alcohol-preferring P rat and animal models of excessive alcohol drinking. Addict Biol. 2006; 11(3-4):270–288. [PubMed: 16961759]

44. Sommer W, Hyytiä P, Kiianmaa K. The alcohol-preferring AA and alcohol-avoiding ANA rats: neurobiology of the regulation of alcohol drinking. Addict Biol. 2006; 11(3-4):289–309. [PubMed: 16961760]

45. Vengeliene V, Siegmund S, Singer MV, Sinclair JD, Li TK, Spanagel R. A comparative study on alcohol-preferring rat lines: effects of deprivation and stress phases on voluntary alcohol intake. Alcohol ClinExp Res. 2003; 27(7):1048–1054.

46. Vengeliene V, Leonardi-Essmann F, Perreau-Lenz S, Gebicke-Haerter P, Drescher K, Gross G, Spanagel R. The dopamine D3 receptor plays an essential role in alcohol-seeking and relapse. FASEB J. 2006; 20(13):2223–2233. [PubMed: 17077299]

47. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004; 5(10):R80. [PubMed: 15461798]

48. Reimers M, Heilig M, Sommer WH. Gene discovery in neuropharmacological and behavioral studies using Affymetrix microarray data. Methods. 2005; 37(3):219–228. [PubMed: 16308151]

49. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000; 25(1):25–29. [PubMed: 10802651]

50. Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. Genet Epidemiol. 2008; 32(3):227–234. [PubMed: 18300295]

51. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol. 2008; 32(4):381–385. [PubMed: 18348202]

52. Nurnberger JI Jr, Foroud T, Flury L, Su J, Meyer ET, Hu K, Crowe R, Edenberg H, Goate A, Bierut L, Reich T, Schuckit M, Reich W. Evidence for a locus on chromosome 1 that influences vulnerability to alcoholism and affective disorder. Am J Psychiatry. 2001; 158(5):718–724. [PubMed: 11329392]

53. Schuckit MA, Edenberg HJ, Kalmijn J, Flury L, Smith TL, Reich T, Bierut L, Goate A, Foroud T. A genome-wide search for genes that relate to a low level of response to alcohol. Alcohol ClinExp Res. 2001; 25(3):323–329.

54. Schuckit MA, Smith TL, Trim R, Kreikebaum S, Hinga B, Allen R. Testing the level of response to alcohol-based model of heavy drinking and alcohol problems in offspring from the San Diego prospective study. J Stud Alcohol Drugs. 2008; 69(4):571–579. [PubMed: 18612573]

55. Baldwin HA, Wall TL, Schuckit MA, Koob GF. Differential effects of ethanol on punished responding in the P and NP rats. Alcohol ClinExp Res. 1991; 15(4):700–704.

56. Moore MS, DeZazzo J, Luk AY, Tully T, Singh CM, Heberlein U. Ethanol intoxication in Drosophila: Genetic and pharmacological evidence for regulation by the cAMP signaling pathway. Cell. 1998; 93(6):997–1007. [PubMed: 9635429]

57. Li TK. Pharmacogenetics of responses to alcohol and genes that influence alcohol drinking. J Stud Alcohol. 2000; 61(1):5–12. [PubMed: 10627090]

58. Heath AC, Martin NG. Genetic differences in psychomotor performance decrement after alcohol: a multivariate analysis. J Stud Alcohol. 1992; 53(3):262–271. [PubMed: 1583905]

59. Madden PA, Heath AC, Starmer GA, Whitfield JB, Martin NG. Alcohol sensitivity and smoking history in men and women. Alcohol ClinExp Res. 1995; 19(5):1111–1120.

60. Martin NG. Twin Studies of alcohol consumption, metabolism and sensitivity. Australian Drug and Alcohol Review. 1988; (7):9–12.

61. Begleiter H, Porjesz B, Reich T, Edenberg HJ, Goate A, Blangero J, Almasy L, Foroud T, Van Eerdewegh P, Polich J, Rohrbaugh J, Kuperman S, Bauer LO, O'Connor SJ, Chorlian DB, Li TK, Conneally PM, Hesselbrock V, Rice JP, Schuckit MA, Cloninger R, Nurnberger J Jr, Crowe R, Bloom FE. Quantitative trait loci analysis of human event-related brain potentials: P3 voltage. Electroencephalogr Clin Neurophysiol. 1998; 108(3):244–250. [PubMed: 9607513]

62. Porjesz B, Begleiter H, Wang K, Almasy L, Chorlian DB, Stimus AT, Kuperman S, O'Connor SJ, Rohrbaugh J, Bauer LO, Edenberg HJ, Goate A, Rice JP, Reich T. Linkage and linkage disequilibrium mapping of ERP and EEG phenotypes. Biol Psychol. 2002; 61(1-2):229–248. [PubMed: 12385677]

63. Polich J, Pollock VE, Bloom FE. Meta-analysis of P300 amplitude from males at risk for alcoholism. Psychol Bull. 1994; 115(1):55–73. [PubMed: 8310100]

64. Benegal V, Jain S, Subbukrishna DK, Channabasavanna SM. P300 amplitudes vary inversely with continuum of risk in first degree male relatives of alcoholics. Psychiatr Genet. 1995; 5(4):149–156. [PubMed: 8750356]

65. Hill SY, Shen S, Zezza N, Hoffman EK, Perlin M, Allan W. A genome wide search for alcoholism susceptibility genes. Am J Med Genet B Neuropsychiatr Genet. 2004; 128B(1):102–113. [PubMed: 15211641]

66. Tellegen, A.; Waller, NG. Exploring personality through test construction: development of the multidimensional personality questionnnaire. Minneapolis: University of Minnesota Press; 1982.

67. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet. 1995; 11(3):241–247. [PubMed: 7581446]

68. Nestel PJ, Hirsch EZ. Mechanism of alcohol-induced hypertriglyceridemia. J Lab Clin Med. 1965; 66(3):357–365. [PubMed: 5835967]

69. Sun GY, Rush A, Chin PC, Gorka C, Lahiri S, Wood WG. Serum lipids and acyl group composition of alcoholic patients. Alcohol. 1988 Mar-Apr;5(2):153–7. [PubMed: 3395462]

70. Hamid A, Kaur J. Decreased expression of transporters reduces folate uptake across renal absorptive surfaces in experimental alcoholism. J Membr Biol. 2007; 220(1-3):69–77. [PubMed: 18008023]

71. Das AK, Uhler MD, Hajra AK. Molecular cloning and expression of mammalian peroxisomal trans-2-enoyl-coenzyme A reductase cDNAs. J Biol Chem. 2000; 275(32):24333–24340. [PubMed: 10811639]

72. Prescott CA, Sullivan PF, Kuo PH, Webb BT, Vittum J, Patterson DG, Thiselton DL, Myers JM, Devitt M, Halberstadt LJ, Robinson VP, Neale MC, van den Oord EJ, Walsh D, Riley BP, Kendler KS. Genomewide linkage study in the Irish affected sib pair study of alcohol dependence: evidence for a susceptibility region for symptoms of alcohol dependence on chromosome 4. Mol Psychiatry. 2006; 11(6):603–611. [PubMed: 16534506]

73. Uhl GR, Drgon T, Liu QR, Johnson C, Walther D, Komiyama T, Harano M, Sekine Y, Inada T, Ozaki N, Iyo M, Iwata N, Yamada M, Sora I, Chen CK, Liu HC, Ujike H, Lin SK. Genome-wide

association for methamphetamine dependence: convergent results from 2 samples. Arch Gen Psychiatry. 2008; 65(3):345–355. [PubMed: 18316681]

74. Ioannidis JP, Patsopoulos NA, Evangelou E. Heterogeneity in meta-analyses of genome-wide association investigations. PLoS ONE. 2007; 2(9):e841. [PubMed: 17786212]

75. Chen WJ, Loh EW, Hsu YP, Chen CC, Yu JM, Cheng AT. Alcohol-metabolising genes and alcoholism among Taiwanese Han men: independent effect of ADH2, ADH3 and ALDH2. Br J Psychiatry. 1996; 168(6):762–767. [PubMed: 8773821]

76. Chen CC, Lu RB, Chen YC, Wang MF, Chang YC, Li TK, Yin SJ. Interaction between the functional polymorphisms of the alcohol-metabolism genes in protection against alcoholism. Am J Hum Genet. 1999; 65(3):795–807. [PubMed: 10441588]

77. Osier M, Pakstis AJ, Kidd JR, Lee JF, Yin SJ, Ko HC, Edenberg HJ, Lu RB, Kidd KK. Linkage disequilibrium at the ADH2 and ADH3 loci and risk of alcoholism. Am J Hum Genet. 1999; 64(4):1147–1157. [PubMed: 10090900]

78. Ehlers CL, Gilder DA, Harris L, Carr L. Association of the ADH2*3 allele with a negative family history of alcoholism in African American young adults. Alcohol ClinExp Res. 2001; 25(12):1773–1777.

79. Ehlers CL, Montane-Jaime K, Moore S, Shafe S, Joseph R, Carr LG. Association of the ADHIB*3 allele with alcohol-related phenotypes in Trinidad. Alcohol ClinExp Res. 2007; 31(2):216–20.

80. Whitfield JB, Nightingale BN, Bucholz KK, Madden PA, Heath AC, Martin NG. ADH genotypes and alcohol use and dependence in Europeans. Alcohol Clin Exp Res. 1998; 22(7):1463–1469. [PubMed: 9802529]

81. Borràs E, Coutelle C, Rosell A, Fernández-Muixi F, Broch M, Crosas B, Hjelmqvist L, Lorenzo A, Gutiérrez C, Santos M, Szczepanek M, Heilig M, Quattrocchi P, Farrés J, Vidal F, Richart C, Mach T, Bogdal J, Jörnvall H, Seitz HK, Couzigou P, Parés X. Genetic polymorphism of alcohol dehydrogenase in europeans: the ADH2*2 allele decreases the risk for alcoholism and is associated with ADH3*1. Hepatology. 2000; 31(4):984–989. [PubMed: 10733556]

82. Edenberg HJ, Xuei X, Chen HJ, Tian H, Wetherill LF, Dick DM, Almasy L, Bierut L, Bucholz KK, Goate A, Hesselbrock V, Kuperman S, Nurnberger J, Porjesz B, Rice J, Schuckit M, Tischfield J, Begleiter H, Foroud T. Association of alcohol dehydrogenase genes with alcohol dependence: a comprehensive analysis. Hum Mol Genet. 2006; 15(9):1539–1549. [PubMed: 16571603]

83. Luo X, Kranzler HR, Zuo L, Wang S, Schork NJ, Gelernter J. Diplotype trend regression analysis of the ADH gene cluster and the ALDH2 gene: multiple significant associations with alcohol dependence. Am J Hum Genet. 2006; 78(6):973–987. [PubMed: 16685648]

84. Kuo PH, Kalsi G, Prescott CA, Hodgkinson CA, Goldman D, van den Oord EJ, Alexander J, Jiang C, Sullivan PF, Patterson DG, Walsh D, Kendler KS, Riley BP. Association of ADH and ALDH genes with alcohol dependence in the Irish Affected Sib Pair Study of alcohol dependence (IASPSAD) sample. Alcohol Clin Exp Res. 2008; 32(5):785–795. [PubMed: 18331377]

85. Zintzaras E, Stefanidis I, Santos M, Vidal F. Do alcohol-metabolizing enzyme gene polymorphisms increase the risk of alcoholism and alcoholic liver disease? Hepatology. 2006; 43(2):352–361. [PubMed: 16440362]

86. Neale BM, Sham PC, Purcell S, Banaschewski T, Buitelaar J, Franke B, Sonuga-Barke E, Ebstein R, Eisenberg J, Mulligan A, Gill M, Manor I, Miranda A, Mulas F, Oades RD, Roeyers H, Rothenberger A, Sergeant J, Steinhausen HC, Taylor E, Thompson M, Chen W, Zhou K, Asherson P, Faraone SV. Population differences in the International Multi-Centre ADHD Gene Project. Genet Epidemiol. 2008; 32(2):98–107. [PubMed: 17868146]

87. Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, Stephens M, Bustamante CD. Genes mirror geography within Europe. Nature. 456:98–101. [PubMed: 18758442]

88. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38(8):904–909. [PubMed: 16862161]
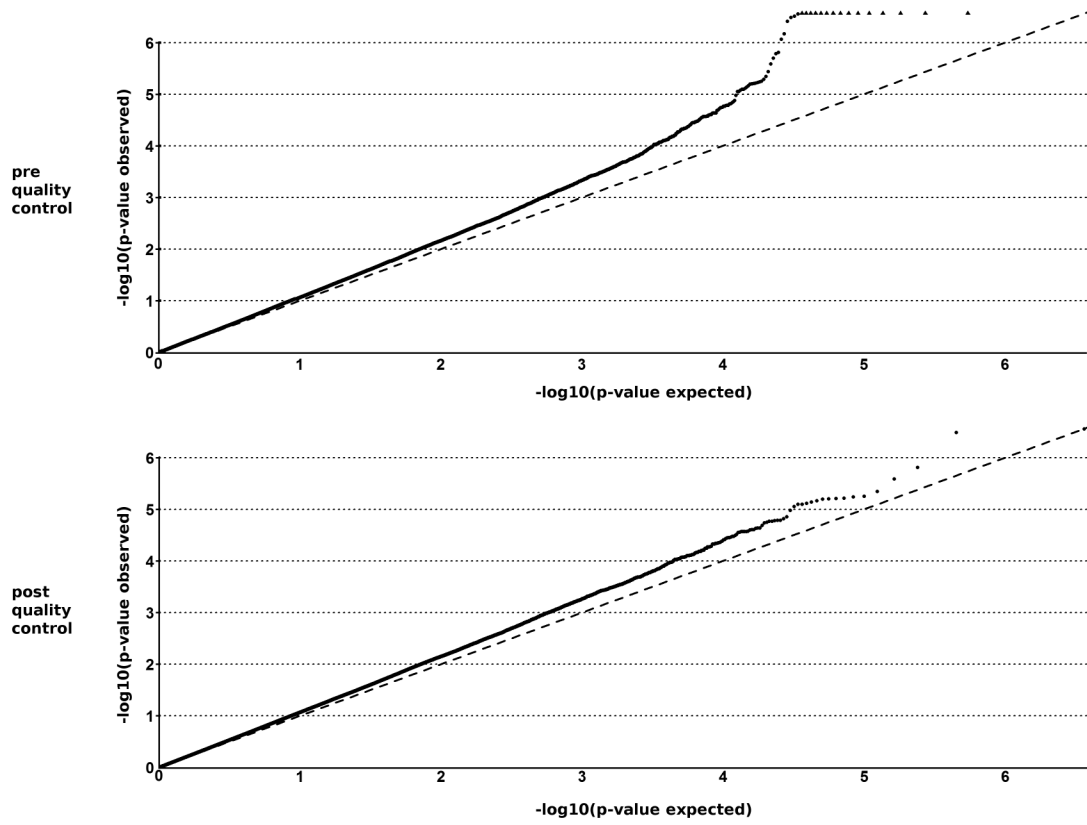
**Figure 1.**
Pre and post quality control QQ plot of Armitage's trend test p-values of autosomal SNPs.
SNPs at which the test statistic exceeds 30, are represented by triangles at the top of the plot.
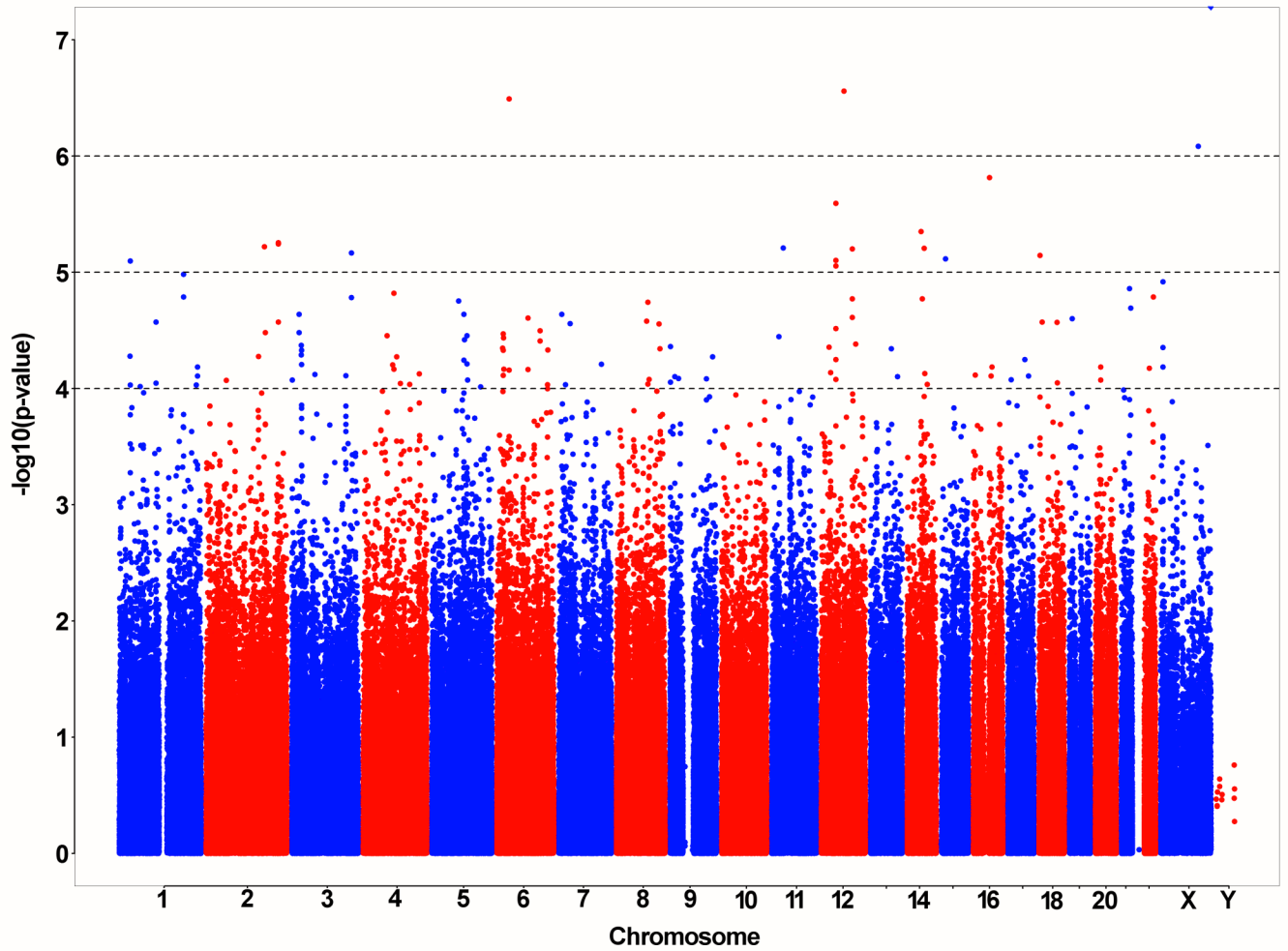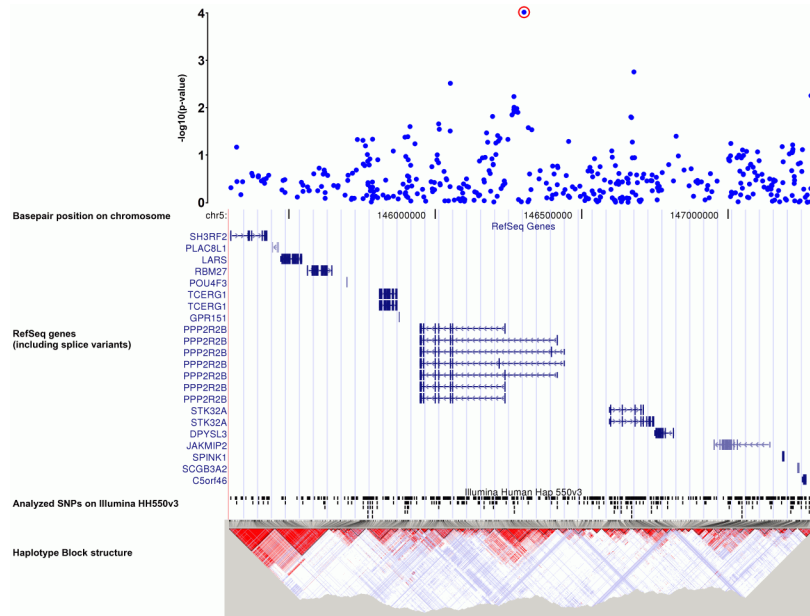
**Figure 2.**
Genome-wide overview of GWAS findings. SNPs at which the test statistic exceeds 30, are represented by triangles at the top of the plot.
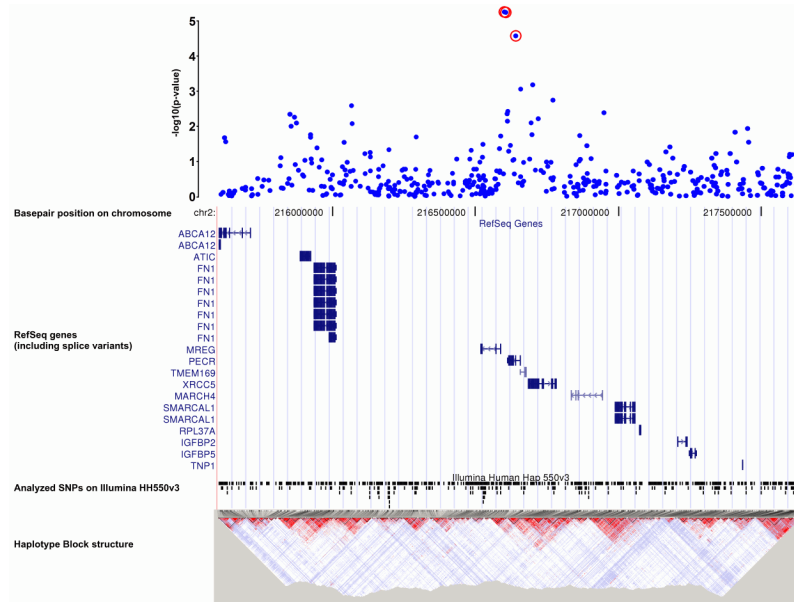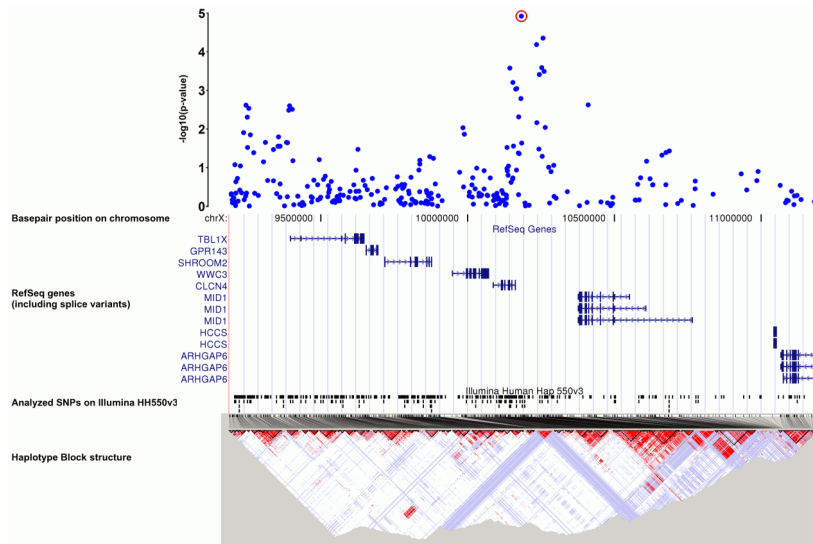
**Figure 3.**
Illustration of p-values in a +/− 1Mb interval around the most significant findings at 2q35: rs1344694, rs7590720 and rs705648. Gene annotations from UCSC RefSeq genes, LD-maps from HapMap.

**Figure 4.**
Illustration of p-values in a +/− 1Mb interval around the most significant finding rs1864982 at 5q32.Gene annotations from UCSC RefSeq genes, LD-maps from HapMap.

**Figure 5.**
Illustration of p-values in a +/− 1Mb interval around the most significant finding rs12388359 at Xp22.2. Gene annotations from UCSC RefSeq genes, LD-maps from HapMap.

**Table 1**

**SNPs confirmed in the follow-up study – location according to chromosomal bands and gene annotation**

| SNP | Chromosomal band | Genes[a] |
|---|---|---|
| rs1344694 | 2q35 | |
| rs7590720 | 2q35 | |
| rs705648 | 2q35 | Peroxisomal trans-2-enoyl-CoA reductase *(PECR)* |
| rs1614972[b] | 4q23 | Alcohol dehydrogenase 1C (class I), gamma polypeptide *(ADH1C)* |
| rs13362120 | 5q15 | Calpastatin *(CAST)* |
| rs13160562 | 5q15 | Endoplasmic reticulum aminopeptidase 1 *(ERAP1)*; calpastatin *(CAST)* |
| rs1864982 | 5q32 | Protein phosphatase 2 (formerly 2A), regulatory subunit B, beta isoform *(PPP2R2B)* |
| rs6902771 | 6q25.1 | Estrogen receptor 1 *(ESR1)* |
| rs729302 | 7q32.1 | |
| rs13273672[b] | 8p23.1 | GATA binding protein 4 *(GATA4)* |
| rs1487814 | 11p14.3 | |
| rs7138291 | 12q22 | Coiled-coil domain containing 41 *(CCDC41)* |
| rs36563 | 14q24.2 | |
| rs11640875[b] | 16q23.3 | Cadherin 13, H-cadherin (heart) *(CDH13)* |
| rs12388359 | Xp22.2 | |

[a] annotation according to dbSNP build 129

[b] selected following strategy of "rodent candidate gene"

**Table 2**

SNPs confirmed in the follow-up study – Comparison of results of GWAS, replication and pooled analysis.

| SNP | Allele A[d] | Allele B[d] | GWA Scan[a] | | | | Replication Sample[b] | | | | Pooled Sample[c] | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | Frq A Ca | Frq A Co | P-value[e] | Odds Ratio (95% CI) | Frq A Ca | Frq A Co | P-value[e] | Odds Ratio (95% CI) | Frq A Ca | Frq A Co | P-value[e] | Odds Ratio (95% CI) |
| **rs1344694** | G | T | 0.603 | 0.685 | 5.57E-6 | 0.70 (0.60–0.81) | 0.613 | 0.659 | 3.36E-3 | 0.82 (0.72–0.93) | 0.610 | 0.674 | 1.69E-8 | 0.76 (0.69–0.83) |
| **rs7590720** | A | G | 0.645 | 0.724 | 5.70E-6 | 0.69 (0.59–0.81) | 0.661 | 0.713 | 6.68E-4 | 0.79 (0.68–0.90) | 0.656 | 0.719 | 9.72E-9 | 0.74 (0.67–0.82) |
| **rs705648** | C | T | 0.313 | 0.244 | 2.68E-5 | 1.41 (1.20–1.67) | 0.292 | 0.256 | 1.24E-2 | 1.19 (1.04–1.37) | 0.299 | 0.249 | 1.78E-6 | 1.29 (1.16–1.43) |
| **rs1614972** | C | T | 0.754 | 0.690 | 2.84E-4 | 1.37 (1.16–1.61) | 0.721 | 0.690 | 3.62E-2 | 1.16 (1.01–1.33) | 0.732 | 0.690 | 1.41E-4 | 1.23 (1.10–1.36) |
| **rs13362120** | C | T | 0.258 | 0.329 | 5.72E-5 | 0.71 (0.60–0.84) | 0.296 | 0.333 | 1.25E-2 | 0.84 (0.73–0.96) | 0.283 | 0.331 | 1.85E-5 | 0.80 (0.72–0.88) |
| **rs13160562** | A | G | 0.252 | 0.326 | 2.31E-5 | 0.70 (0.59–0.83) | 0.283 | 0.316 | 2.47E-2 | 0.85 (0.75–0.98) | 0.273 | 0.322 | 7.09E-6 | 0.79 (0.71–0.88) |
| **rs1864982** | A | C | 0.181 | 0.129 | 9.71E-5 | 1.49 (1.22–1.82) | 0.163 | 0.130 | 4.47E-3 | 1.30 (1.09–1.56) | 0.169 | 0.130 | 3.46E-6 | 1.36 (1.20–1.55) |
| **rs6902771** | C | T | 0.591 | 0.517 | 9.31E-5 | 1.35 (1.16–1.56) | 0.552 | 0.505 | 3.15E-3 | 1.20 (1.06–1.37) | 0.565 | 0.512 | 8.30E-6 | 1.24 (1.13–1.36) |
| **rs729302** | A | C | 0.599 | 0.671 | 6.19E-5 | 0.73 (0.63–0.85) | 0.641 | 0.677 | 1.84E-2 | 0.85 (0.75–0.97) | 0.627 | 0.673 | 4.02E-5 | 0.82 (0.74–0.90) |
| **rs13273672** | C | T | 0.364 | 0.312 | 2.18E-3 | 1.27 (1.09–1.47) | 0.342 | 0.311 | 3.61E-2 | 1.15 (1.01–1.32) | 0.350 | 0.312 | 4.75E-4 | 1.19 (1.08–1.31) |
| **rs1487814** | A | G | 0.589 | 0.511 | 3.58E-5 | 1.37 (1.18–1.59) | 0.544 | 0.508 | 2.48E-2 | 1.15 (1.02–1.30) | 0.559 | 0.510 | 3.49E-5 | 1.22 (1.11–1.34) |
| **rs7138291** | C | T | 0.142 | 0.093 | 2.46E-5 | 1.61 (1.28–2.00) | 0.115 | 0.095 | 4.10E-2 | 1.23 (1.01–1.54) | 0.124 | 0.094 | 3.68E-5 | 1.36 (1.18–1.58) |
| **rs36563** | A | C | 0.210 | 0.147 | 6.21E-6 | 1.54 (1.28–1.85) | 0.184 | 0.159 | 3.83E-2 | 1.19 (1.01–1.41) | 0.192 | 0.152 | 4.57E-6 | 1.33 (1.18–1.50) |
| **rs11640875** | A | G | 0.389 | 0.326 | 5.25E-4 | 1.32 (1.12–1.54) | 0.372 | 0.332 | 9.85E-3 | 1.19 (1.04–1.35) | 0.377 | 0.329 | 1.84E-5 | 1.24 (1.12–1.36) |
| **rs12388359** | G | T | 0.811 | 0.900 | 1.21E-5 | 0.47 (0.34–0.67) | 0.840 | 0.879 | 1.25E-2 | 0.72 (0.56–0.93) | 0.831 | 0.888 | 3.57E-6 | 0.62 (0.50–0.76) |

[a] Analysis based on 476 cases and 1358 controls

[b] Analysis based on 984 cases and 974 controls

[c] Analysis based on 1460 cases and 2332 controls

[d] Allele names are assigned in alphabetical order

[e] Armitage test for trend for autosomal markers, allele-based chi$^2$-Test for x-chromosomal marker