npg

XRCC5 as a Risk Gene for Alcohol Dependence: Evidence from a Genome-Wide Gene-Set-Based Analysis and Follow-up Studies in *Drosophila* and Humans

Dilafruz Juraeva[1,24], **Jens Treutlein**[2,24], **Henrike Scholz**[3,24], **Josef Frank**[2], **Franziska Degenhardt**[4,5], **Sven Cichon**[6], **Monika Ridinger**[7], **Manuel Mattheisen**[8], **Stephanie H Witt**[2], **Maren Lang**[2], **Wolfgang H Sommer**[9], **Per Hoffmann**[6], **Stefan Herms**[6], **Norbert Wodarz**[7], **Michael Soyka**[10,11], **Peter Zill**[11], **Wolfgang Maier**[12], **Elisabeth Jünger**[13], **Wolfgang Gaebel**[14], **Norbert Dahmen**[15], **Norbert Scherbaum**[16], **Christine Schmäl**[2], **Michael Steffens**[17], **Susanne Lucae**[18], **Marcus Ising**[19], **Michael N Smolka**[13], **Ulrich S Zimmermann**[13], **Bertram Müller-Myhsok**[20,21,22], **Markus M Nöthen**[4,5], **Karl Mann**[23], **Falk Kiefer**[23], **Rainer Spanagel**[9,24], **Benedikt Brors**[1,24] and **Marcella Rietschel**[*,2,24]

[1]Division of Theoretical Bioinformatics, German Cancer Research Center, Heidelberg, Germany; [2]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany; [3]Department of Animal Physiology, University of Cologne, Cologne, Germany; [4]Institute of Human Genetics, University of Bonn, Bonn, Germany; [5]Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany; [6]Department of Biomedicine, University Hospital Basel, Basel, Switzerland; [7]Department of Psychiatry, University of Regensburg, Regensburg, Germany; [8]Department of Biomedicine, University of Aarhus, Denmark; [9]Institute of Psychopharmacology, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany; [10]Private Hospital Meiringen, Meiringen, Switzerland; [11]Department of Psychiatry, University of Munich, Munich, Germany; [12]Department of Psychiatry, University of Bonn, Bonn, Germany; [13]Department of Psychiatry and Psychotherapy, University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden; [14]Department of Psychiatry and Psychotherapy, University of Düsseldorf, Düsseldorf, Germany; [15]Department of Psychiatry, University of Mainz, Mainz, Germany; [16]Addiction Research Group at the Department of Psychiatry and Psychotherapy, University of Duisburg-Essen, Essen, Germany; [17]Division of Research, Federal Institute for Drugs and Medical Devices, Bonn, Germany; [18]Department of Psychiatric Pharmacogenetics, Max-Planck-Institute of Psychiatry, München, Germany; [19]Department of Molecular Psychology, Max-Planck-Institute of Psychiatry, München, Germany; [20]Department of Statistical Genetics, Max-Planck-Institute of Psychiatry, München, Germany; [21]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany; [22]Institute of Translational Medicine Liverpool, University of Liverpool, Liverpool, UK; [23]Department of Addictive Behavior and Addiction Medicine, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany

Genetic factors have as large role as environmental factors in the etiology of alcohol dependence (AD). Although genome-wide association studies (GWAS) enable systematic searches for loci not hitherto implicated in the etiology of AD, many true findings may be missed owing to correction for multiple testing. The aim of the present study was to circumvent this limitation by searching for biological system-level differences, and then following up these findings in humans and animals. Gene-set-based analysis of GWAS data from 1333 cases and 2168 controls identified 19 significantly associated gene-sets, of which 5 could be replicated in an independent sample. Clustered in these gene-sets were novel and previously identified susceptibility genes. The most frequently present gene, ie in 6 out of 19 gene-sets, was X-ray repair complementing defective repair in Chinese hamster cells 5 (*XRCC5*). Previous human and animal studies have implicated *XRCC5* in alcohol sensitivity. This phenotype is inversely correlated with the development of AD, presumably as more alcohol is required to achieve the desired effects. In the present study, the functional role of *XRCC5* in AD was further validated in animals and humans. *Drosophila* mutants with reduced function of *Ku80*—the homolog of mammalian *XRCC5*—due to RNAi silencing showed reduced sensitivity to ethanol. In humans with free access to intravenous ethanol self-administration in the laboratory, the maximum achieved blood alcohol concentration was influenced in an allele-dose-dependent manner by genetic variation in *XRCC5*. In conclusion, our convergent approach identified new candidates and generated independent evidence for the involvement of *XRCC5* in alcohol dependence.
Neuropsychopharmacology (2015) **40**, 361–371; doi:10.1038/npp.2014.178; published online 13 August 2014

INTRODUCTION

Formal genetic studies show that the heritability of alcohol dependence (AD) is around 40–60% (Prescott and Kendler, 1999; Sullivan *et al*, 2012). Candidate and genome-wide association studies (GWAS) have suggested numerous candidates for AD and alcohol consumption. The most consistent associations have been reported for: (i) chromosome 4q22/4q23 in/near the alcohol dehydrogenase (*ADH*) gene cluster; and (ii) in/near chromosome 12q24, a region which harbors the aldehyde dehydrogenase 2 (*ALDH2*) gene. Whereas candidate studies rely on prior knowledge, GWAS allow systematic screening of the whole genome with hundreds of thousands to millions of genetic variants, thus facilitating the identification of genes in novel biological contexts. Owing to correction for multiple testing, the probability of identifying genome-wide significant associations hinges on the sample size as well as on the degree of heritability and heterogeneity (Sullivan *et al*, 2012). Genome-wide significant association findings for AD have been detected in: (i) the chr4q22/4q23 region in/near the *ADH* genes, *PDLIM5, METAP1,* and *LOC100507053* (Frank *et al*, 2012; Gelernter *et al*, 2013a; Park *et al*, 2013); (ii) at chr1p35 in *SERINC2* (Zuo *et al*, 2013); (iii) at chr2p16 near *MTIF2, CCDC88A,* and *PRORSD1P* (Gelernter *et al*, 2013a); (iv) at chr2q21 near *DARS* and *CXCR4* (Gelernter *et al*, 2013a); (v) at chr2q35 near *PECR* (Treutlein *et al*, 2009); (vi) at chr5p15 (Gelernter *et al*, 2013a); at chr9p13 (Gelernter *et al*, 2013a); at chr12q24 in *ALDH2* (Quillen *et al*, 2014); (vii) at chr13q32 in *NALCN* (Wetherill *et al*, 2013); and (viii) at chr19p13 in *LOC100131094* and *DPP9* (Gelernter *et al*, 2013a). GWAS of alcohol consumption have identified significant variants at: (i) chr12q24 in/near *ALDH2, C12orf51, CCDC63, MYL2, OAS3, CUX2* (alias *CUTL2*), and *BRAP* (Baik *et al*, 2011; Takeuchi *et al*, 2011); (ii) chr3p24 near *SGOL1* (Pan *et al*, 2013); (iii) chr4q23 in *ADH1B* (McKay *et al*, 2011); and (iv) 7q11 in *AUTS2* (Schumann *et al*, 2011). With the exception of the *ADH* and *ALDH2* genes, none of the variants identified for AD to date exert strong effects, which is consistent with the hypothesis of a polygenic contribution.

In accordance with the latter hypothesis, polygenic score analyses of European GWAS data have shown that many common variants—most of which do not exceed the threshold for genome-wide significance after correction for multiple testing—contribute to the risk of AD (Frank *et al*, 2012). It can therefore be assumed that many risk factors at the level of individual genes still await identification. A promising approach that circumvents extreme correction for multiple testing is to analyze the aggregated contribution of variants in functionally related gene groups under the assumption that these gene groups contain a large number of variants with a disruptive influence on gene function. Specific gene groups can be defined by searching for specific gene-networks (eg grouping genes whose protein products interact (Jia *et al*, 2011)), or by relying on *a priori* information concerning gene-pathways. Recent studies of such system-level approaches have identified new genes for alcohol- (Han *et al*, 2013), cocaine- (Gelernter *et al*, 2013c), and opioid dependence (Gelernter *et al*, 2013b).

The aim of the present study was to identify as yet unknown genes with an involvement in AD using both the gene-set-based analysis of data from a previous GWAS of AD from the German population (Frank *et al*, 2012), and the Global test method (Deelen *et al*, 2013). The most convincing finding, *XRCC5,* was followed up in a functional genetic analysis of its homolog in the fruit fly *Drosophila melanogaster.* The phenotype alcohol sensitivity (Scholz, 2009) was investigated, as this *Drosophila* phenotype corresponds to the human phenotype 'level of response'. The level of response to alcohol refers to the psychological and physiological response of an individual to an alcohol challenge. A previous study reported an association between low level of response to alcohol in humans and an increased risk for AD (Schuckit, 1994). As low responders are likely to consume more alcohol in order to experience the desired pleasurable feelings, we performed a follow-up human genetic association study of *XRCC5* and free access to intravenous ethanol in the laboratory-setting (Zimmermann *et al*, 2009).

MATERIALS AND METHODS

Genome-Wide Association Study

Ethics statement. All participants provided written informed consent following a detailed explanation of the study, and all data were anonymized prior to analysis. The study was approved by the ethics committees of the German Universities of Heidelberg, Bonn, Dresden, Düsseldorf, Essen, Cologne, Mainz, Munich, and Regensburg. All research procedures were conducted in accordance with the Declaration of Helsinki.

GWAS and replication study. The case and control samples from the German GWAS, and the COGA and SAGE GWAS used to replicate significant gene-sets, are described in detail elsewhere (Frank *et al*, 2012, Edenberg *et al*, 2010, Bierut *et al*, 2010; for details, see Supplementary Materials and Methods).

GWAS data set. Briefly, genome-wide data for 1333 cases and 2168 controls were available for analysis after stringent quality control. The quality control criteria were a sample call rate (CR) of $\geqslant 0.98$, and conformity between reported sex and genotypic sex. In the case of duplicates or cryptic relatedness (identity by state across autosomal markers $\geqslant 1.6$), the sample with the lower CR was removed. Outliers were identified using principal component analysis and removed. Only those variants that were present on all of the applied genotyping platforms were included in the analysis. To remain in the data set, single nucleotide polymorphisms (SNPs) had to meet the following criteria: CR $\geqslant 0.98$; minor allele frequency $\geqslant 0.01$; conformity with Hardy–Weinberg equilibrium (HWE; $P \geqslant 1E\text{–}6$) in any subsample; HWE conformity across all samples; and no significant difference in allele frequency between subsamples for controls (Frank *et al*, 2012). Owing to German regulations concerning informed consent and data protection, the data cannot be made available via publicly accessible databases. However, the data have been made available to the Psychiatric Genomics Consortium on AD, which is being established at the time of writing (http://pgc.unc.edu).

Gene-set-based analysis. Gene-set-based analysis was performed using these GWAS data.The association model

used by Frank *et al* (2012) was applied. Briefly, logistic regression was performed, using AD diagnosis as the response variable and the number of minor alleles at considered loci as predictor variables. The principal component analysis in Frank *et al* (2012) revealed strong ethnic homogeneity across the sample. Therefore in the present analyses, no correction for population stratification was performed. Gene-set descriptions were retrieved from the following gene-set collections: Kyoto Encyclopedia of Genes and Genomes (dbKEGG, http://www.genome.jp/kegg/, R package *KEGG.db* version 2.5.0); Reactome (dbRC, http://www.reactome.org/, R package *reactome.db* version 1.44.0); Gene Ontology (dbGO, www.geneontology.org, R package *GO.db* version 2.5.0); Biocarta (dbBC); microRNA targets (dbMIR); transcription factor targets (dbTFT); and positional gene-sets (dbPOS). The latter four gene-sets were retrieved via MSigDB (http://www.broadinstitute.org/gsea/msigdb/, version 3.0).

Gene-sets with a minimum of 5 and a maximum of 200 genes were retained. To reduce multicollinearity in the GWAS data, a pruned SNP set was obtained using a variance inflation factor (VIF = 10) implemented in PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/). Linkage disequilibrium (LD)-based pruning was applied to the raw data using the complete SNP set. After LD-based pruning, 100471 SNPs (from a total of 462775) were considered for the gene-set analysis. To account for important regulatory regions, SNPs were assigned to a gene if the variant was located within the genomic sequence or within 20 kb of the 5′ and 3′ ends of the first and last exons (Veyrieras *et al*, 2008). If a SNP was within a region shared by more than one gene, the SNP was assigned to all of the respective genes. These SNPs were analyzed within the context of 10367 gene-sets (dbBC 217; dbCGP 1817; dbGO 6427; dbKEGG 215; dbMIR 177; dbPOS 288; dbRC 770; dbTFT 456).

Global Test. For the gene-set-based analysis, R package globaltest version 5.12.0 was applied. Details of the Global test are provided elsewhere (Goeman *et al*, 2004; Juraeva *et al*, 2014, for details, see Supplementary Materials and Methods). The contribution of each SNP to the gene-set association score was calculated using the component Global Test. SNPs with a component Global Test *P*-value of $\leqslant$1E-3 are reported as 'top SNPs'. A SNP was classified as a significant contributor to a gene-set association score if its component Global Test *P*-value was $\leqslant$5E-2. A complete list of SNPs contributing to a gene-set association score are provided in Supplementary Table S1. For the top genes in the merged COGA/SAGE replication data set, single SNP analysis and gene-wide analysis were performed. The gene-wide analysis was performed using VEGAS (http://gump.qimr.edu.au/VEGAS/).

Drosophila Study

The most promising finding, ie *XRCC5,* was subjected to gene-targeted functional genetic analysis in an invertebrate model of AD (Scholz, 2009). Homologs of *XRCC5* exist across a wide range of organisms (Altenhoff *et al*, 2013; Li *et al*, 2006). Initial sensitivity to alcohol was used as the phenotype of interest in determining genetic factors that might influence the development of AD. To determine whether altered *Ku80* function interferes with ethanol-induced behaviors, flies with

altered *Ku80* function were tested for changes in ethanol sensitivity using an inebriometer, an assay that measures the effect of ethanol intoxication on postural control (Hoffmann and Cohan, 1987). Briefly, the assay consists of an ethanol-vapor-filled 1.22-meter-long column. A population of 3–5-day-old flies ($n = 100$) is inserted into the top of the device. Over time, the flies lose their balance and fall through the column. At the bottom, the flies are counted by the passing through of a light beam. The time required for a population to elute from the column is defined as mean elution time and is around 20 min for a control population. To reduce *Ku80* function, a RNAi hairpin construct of the *Ku80* gene (dsKu80[JF02790]) was expressed under the control of the UAS/GAL4 system. These experiments involved a population of female flies in which an UAS-Ku80-RNAi transgene was expressed under the control of the GAL4 neuronal driver Appl-GAL4. The latter normally expresses transgenes in all pan-neural cells. To control for the putative effects of the transposon insertion sites and the putative effects of the transgenes alone, heterozygote flies carrying only one copy of the transgene were used as controls. To reduce the influence of modifiers in the genetic background, the transgenic lines were backcrossed for five generations using the w^{1118} stock of the Scholz laboratory. All experiments with *D. melanogaster* were performed in accordance with relevant guidelines and regulations of the German Research Foundation (DFG).

Investigation of Association between Free Access Alcohol Self-Administration and the *XRCC5* Marker rs828701 in Social Drinkers

Subjects. Data sets were available for 85 healthy non-alcohol-dependent social drinkers ($n = 49$ male, $n = 36$ female; mean age: 18.39 ± 0.49 years), all of whom were of German descent. These individuals had been identified from the citizen registry of the German city of Dresden. The majority (76%) were living with their parents, and 77% were in their 13th year of education and planning to attend university. They were phenotyped using a previously described experimental paradigm involving free access to the intravenous self-infusion of ethanol (Zimmermann *et al*, 2009).

Genotyping of XRCC5. Of the four SNPs in *XRCC5,* only rs828701, ie the marker with the most significant *P*-value in this gene, was genotyped. This approach was used as the four top SNPs were intronic, did not constitute expression quantitative trait loci (eQTLs), and had no other reported functional effects. Genotyping was performed on an Applied Biosystems 7900HT Fast Real-Time PCR System (assay ID: C___8839929_10), in accordance with the manufacturer's instructions (Applied Biosystems, Darmstadt, Germany). An ANOVA (function Anova, package car in R) was used to test the effects of rs828701 on maximum achieved blood alcohol concentration (BAC).

RESULTS

Gene-Set Association Analysis

Nineteen gene-sets with a false discovery rate (FDR) of $\leqslant$5E-2 were identified (Table 1, Supplementary Table S1). FDR values ranged from 9.01E-3 to 4.62E-2. Relationships

Table I Associated Gene-Sets of the German GWAS and Replication

Gene-set	FDR (German GWAS)	P (German GWAS)	Variants contributing to gene-set association at $P \leqslant 0.001$ (raw *P*-value in brackets) in the German GWAS	*P* (merged COGA/SAGE GWAS)
(1) dbBC:100021:telomeres telomerase cellular aging and immortality	1.50E-2	6.89E-5	rs4381631_*PRKCA* (4.97E-4), **rs828701_*XRCC5* (2.25E-5), rs828704_*XRCC5* (3.84E-5), rs207938_*XRCC5* (8.71E-4), rs2032765_*XRCC5* (8.99E-4)**	n.s.
(2) dbGO:0019047:provirus integration	4.15E-2	2.99E-5	rs17134462_*PPIA* (2.39E-5), rs4724321_*PPIA*(4,96E-4), **rs828701_*XRCC5* (2.25E-5), rs828704_*XRCC5* (3.84E-5), rs207938_*XRCC5*(8.71E-4), rs2032765_*XRCC5* (8.99E-4)**	n.s.
(3) dbGO:0005095:GTPase inhibitor activity	1.89E-2	3.57E-5	rs4732254_*DGKI* (1.68E-4), rs10488609_*DGKI* (1.97E-4), rs4728409_*DGKI* (4.10E-4), rs1878892_*RHOH* (6.33E-4)	n.s.
(4) dbGO:0070419: nonhomologous end joining complex	2.16E-2	9.33E-5	**rs828701_*XRCC5* (2.25E-5), rs828704_*XRCC5* (3.84E-5), rs207938_*XRCC5* (8.71E-4), rs2032765_*XRCC5* (8.99E-4)**	n.s.
(5) dbGO:0015074:DNA integration	4.15E-2	7.20E-3	rs17134462_*PPIA* (2.39E-5), rs4724321_*PPIA* (4,96E-4), **rs828701_ *XRCC5* (2.25E-5), rs828704_*XRCC5* (3.84E-5), rs207938_*XRCC5* (8.71E-4), rs2032765_*XRCC5* (8.99E-4)**	3.07E-2
(6) dbKEGG:04146:peroxisome	1.43E-2	6.64E-5	rs1344694_*PECR* (1.76E-7), rs17548312_*PECR* (6.94E-5)	n.s.
(7) dbKEGG:04510:focal adhesion	3.30E-2	3.75E-4	rs4076090_*LAMA2* (4.08E-4), rs4381631_*PRKCA* (4.97E-4), rs10487171_*RELN* (3.98E-4), rs10272602_*RELN* (3.99E-4)	3.80E-3
(8) dbKEGG:04670:leukocyte transendothelial migration	3.30E-2	4.60E-4	rs170183_*CLDN14* (6.80E-5), rs4381631_*PRKCA* (4.97E-4), rs1878892_*RHOH* (6.33E-4), rs3769232_*RAPGEF4* (8.81E-4)	3.91E-3
(9) dbKEGG:00010:glycolysis/ gluconeogenesis	3.35E-2	7.42E-4	rs1789891_*ADH1B* (5.48E-8), rs2173201_*ADH1B* (1.66E-4), rs1789891_*ADH1C* (5.48E-8), rs11499823_*ADH1C* (1.82E-7), rs2173201_*ADH1C* (1.66E-4)	n.s.
(10) dbKEGG:04142:lysosome	3.35E-2	8.33E-4	rs366193_*ATP6V0A4* (4.70E-4), rs7641926_*SUMF1* (6.04E-4)	n.s.
(11) dbKEGG:03450: nonhomologous end-joining	3.35E-2	1.09E-3	**rs828701_*XRCC5* (2.25E-5), rs828704_*XRCC5* (3.84E-5), rs2032765_*XRCC5* (8.99E-4)**	n.s.
(12) dbKEGG:01040: biosynthesis of unsaturated fatty acids	3.35E-2	1.09E-3	rs1344694_*PECR* (1.76E-7), rs17548312_*PECR* (6.94E-5)	n.s.
(13) dbMIR:tgtgtga,mir-377	2.40E-2	5.45E-4	rs2744413_*ATXN1* (3.18E-5), rs12196135_*ATXN1* (8.75E-5), rs7738608_*ATXN1* (4.37E-4), rs2282832_*ATXN1* (8.13E-4), rs10089142_*EPB49* (6.46E-4), rs3846714_*FAT2* (6.48E-4), rs2520361_*HDAC9* (6.54E-4), rs7641926_*SUMF1* (6.04E-4), rs16911710_*TEAD1* (5.11E-4)	3.51E-2
(14) dbPOS:chr12q12	1.25E-2	4.32E-5	rs400774_*KRT6B* (1.59E-4), rs555740_*SLC2A13* (8.96E-4)	n.s.
(15) dbPOS:chr5q21	3.87E-2	2.69E-4	rs2116823_*FBXL17* (1.17E-4), rs288193_*FBXL17* (1.30E-4), rs9326736_*RPS20P3* (6.14E-4)	n.s.
(16) dbPOS:chr2q35	4.62E-2	6.31E-4	rs1344694_*PECR* (1.76E-7), rs17548312_*PECR* (6.94E-5), rs705648_*PECR* (2.00E-4), rs2032765_*PKI55* (8.99E-4), rs17548312_*TMEM169* (6.94E-5), rs705648_*TMEM169* (2.00E-4), **rs828701_*XRCC5* (2.25E-5), rs828704_*XRCC5* (3.84E-5), rs207938_*XRCC5* (8.71E-4), rs2032765_*XRCC5* (8.99E-4)**	n.s.
(17) dbPOS:chr7q32	4.62E-2	6.42E-4	rs4732254_*DGKI* (1.68E-4), rs10488609_*DGKI* (1.97E-4), rs4728409_*DGKI* (4.10E-4), rs10954651_*HIPK2* (5.30E-4)	2.39E-2
(18) dbTFT:v$srf_q5_01	9.01E-3	4.87E-5	rs1360567_*ADAMTSL1* (6.13E-4), rs10089142_*FGF17* (6.46E-4), rs1568555_*FOXP1* (5.33E-5), rs9309515_*LRRTM4*,(9.11E-4), rs17236557_*MYO1E* (6.28E-4), rs943617_*SUSD1* (9.20E-4), rs1247421_*SVIL* (6.96E-4)	n.s.
(19) dbTFT:v$foxj2_02	9.01E-3	5.93E-5	rs1360567_*ADAMTSL1* (6.13E-4), rs7126303_*ARNTL* (3.68E-4), rs8067516_*ATP2A3* (5.16E-5), rs1568555_*FOXP1* (5.33E-5), rs2520361_*HDAC9* (6.54E-4), rs6719711_*LRP1B* (1.14E-4), rs11901153_*LRP1B* (1.55E-4), rs3923350_*LRP1B* (7.52E-4), rs3769554_*LTBP1* (6.07E-5), rs16911710_*TEAD1* (5.11E-4), rs6932636_*TRERF1* (4.15E-4)	n.s.

Only variants with the strongest contribution ($P \leqslant$ 1E-3) to gene-set significance are shown ('top SNPs'). Variants located at the *XRCC5* locus are in bold. For variants located within a region shared by more than one gene, only the gene belonging to the respective gene-set is shown. For a complete list of the markers contributing to gene-set associations, see Supplementary Table S1. n.s. = not significant (*P* > 0.05).

between gene-sets, as revealed by SNPs with an individual SNP *P*-value of ⩽5E-2, are shown in Supplementary Figure S1. The 19 gene-sets with FDR ⩽5E-2 included five sets, which characterize processes of DNA integrity and DNA repair, ie 'telomeres telomerase cellular aging and immortality' (set 1); 'provirus integration' (set 2); 'non-homologous end joining complex' (set 4); 'DNA integration' (set 5); and 'nonhomologous end-joining' (set 11). The 19

gene-sets also included: the 'GTPase inhibitor activity' set containing gene products, which prevent enzymatic hydrolysis of guanosin triphosphate (set 3); the 'peroxisome' set containing gene products, which have a role in lipid homeostasis and redox reactions of this organelle (set 6); the 'focal adhesion' set containing gene products, which form specialized structures at contact points between the cell and the extracellular matrix (set 7); the 'leukocyte transendothelial migration' set containing gene products of relevance to the migration of leucocytes between blood and tissue (set 8); the 'glycolysis/gluconeogenesis' set, which encodes enzymes for the synthesis of pyruvate from glucose and for the formation of glucose from noncarbohydrate precursors (set 9); the 'lysosome' set containing the protein configuration of this digestive compartment (set 10); the 'biosynthesis of unsaturated fatty acids' set containing the enzymes required for formation of these compounds (set 12); the 'mir-377' set containing genes that share the microRNA binding motif 'tgtgtga' in the 3'-untranslated region (set 13); the four positional sets 'chr12q12' (set 14), 'chr5q21' (set 15), 'chr2q35' (set 16) and 'chr7q32' (set 17), which contain the genes of the respective cytogenetic bands; and the two transcription factor target sets 'srf' (set 18); and 'foxj2' (set 19), which contain genes that share the respective transcription factor binding site.

Association with four of these gene-sets, ie 'peroxisome' (set 6); 'glycolysis/gluconeogenesis' (set 9); 'biosynthesis of unsaturated fatty acids' (set 12); and 'chr2q35' (set 16), was driven by the following three highly significant association findings: ($P < 1E$-5): rs1789891 between *ADH1B* and *ADH1C* (set 9); rs11499823 near *ADH1C* (set 9); and rs1344694 near *PECR* (sets 6,12,16). The remaining 15 gene-sets were defined exclusively by variants with less significant P-values (Table 1).

The 19 gene-sets were comprised of the 38 most significant contributory genes in the most significant category ($P < 1E$-3) (Table 1). The gene *XRCC5* was present in six of the 19 gene-associated sets (Table 1). These gene-sets were: 'telomeres telomerase cellular aging and immortality'; 'nonhomologous end joining'; 'DNA integration'; 'provirus integration'; 'nonhomologous end joining complex'; and the positional gene-set 'chr2q35'. Nine variants at the *XRCC5* locus contributed with nominal significance to the association of these six gene-sets (Supplementary Table S1). These included four variants in the top category $P \leqslant 1E$-3, ie rs828701, rs828704, rs207938, and rs2032765.

The top *XRCC5* marker rs828701 (risk allele: C), ie the SNP with the most significant P-value among these four SNPs, showed $P = 2.25E$-5 (Table 1). The rs828701 C-allele frequency was 48.6% in cases and 43.5% in controls, respectively.

XRCC5 was present in 65 of the 10 367 analyzed gene-sets. A total of 1737 genes were present in at least 65 (or more) gene-sets. This indicates that the frequent presence of *XRCC5* was not due to overrepresentation, and that *XRCC5* had no greater *a priori* likelihood of appearing in the 19 significantly associated gene-sets, ie that this gene was truly enriched for variants that drive the association of the gene-sets.

In the replication study, the set 'DNA integration' (Gene Ontology: 0015074), which contains the *XRCC5* gene, achieved a nominally significant P-value of 3.07E-2. In addition, four further sets achieved a nominally significant P-value of $\leqslant 5E$-2, ie focal adhesion (KEGG: 04510) $P = 3.80E$-3; leucocyte transendothelial migration (KEGG: 04670) $P = 3.91E$-3; chr7q32 (dbPOS-mSIGdb) $P = 2.39E$-2; and MIR-377 (dbMIR-mSIGdb) $P = 3.51E$-2 (Table 1). Thus a nominally significant result was obtained for 5 of the 19 gene-sets tested, which is a significantly higher number than would be expected by chance ($P = 0.002$). For the 38 top genes in the merged COGA/SAGE data set, single-marker analyses in the COGA/SAGE data set generated no significant finding ($P > 5E$-2). Gene-wide analysis using VEGAS revealed nominally significant P-values for *CLDN14* and *SLC2A13*.

Functional Genetic Study in Ku80, the Drosophila Homolog of XRCC5

The one-to-one ortholog of the human gene *XRCC5* in *Drosophila* is the *Ku80* gene (Altenhoff et al, 2013; Li et al, 2006), (Supplementary Figure S2). Experiments in *Drosophila* showed that initial sensitivity to ethanol, as measured by the effect of ethanol intoxication on postural control, was significantly reduced ($P < 0.05$) in *Ku80* mutants compared with controls. This finding supports the hypothesis that *Ku80* has a functional role in the regulation of ethanol sensitivity (Figure 1).

Human Genetic Study of Free Access to Alcohol Self-Administration

The *XRCC5* rs828701 genotypes were in HWE ($P > 0.05$). Their distribution was (i) CC = 8, CT = 24, TT = 17 in the 49 males; and (ii) CC = 7, CT = 19, TT = 10 in the 36 females. The C-allele frequency in non-alcohol-dependent social drinkers was 44.9%, and no sex difference was observed (χ^2-test $P = 0.78$). The rs828701 variant had a robust effect on the self- administration of ethanol (Figure 2). An ANOVA model was computed to predict maximum achieved BACs from genotype. This revealed a significant variation in maximum BACs across genotypes, with TT > CT > CC ($F(2,82) = 3.6$, $P = 0.03$). Sex did not interact with genotype when entered as a covariate into this model. These results show that the T allele was associated with high alcohol consumption in the present experiments.

DISCUSSION

The present results implicate 19 gene-sets in AD susceptibility, all of which encode functionally diverse pathways. In the replication data set, five gene-sets achieved nominally significant P-values ranging from 2.39E-2 to 3.80E-3, ie DNA integration, focal adhesion, leucocyte transendothelial migration, chr7q32, and microRNA mir-377. The replication of this number of gene-sets in an independent sample is significantly higher than would be expected by chance ($P = 0.002$). This finding is consistent with the hypothesis that multiple genes and genetic variants with low effect sizes contribute to AD (Frank et al, 2012), and underlines the validity of such a systematic exploitation of GWAS data with respect to the identification of vulnerability genes.

On the level of gene-sets/biological processes, one of the present findings has been implicated in previous studies of
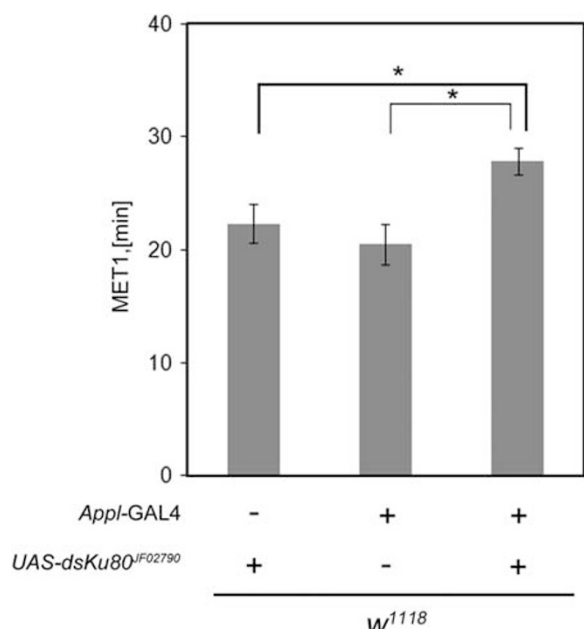
Figure 1 Effect of alcohol on postural control in *Drosophila* Ku80 mutants. The expression of the UAS-dsKu80^{JF02790}: RNAi hairpin construct of the Ku80 gene under the control of the pan-neural GAL4 driver *Appl*-GAL4 significantly reduced ethanol sensitivity in the flies. To control for transgene insertion effects, the transheterozygote *Appl*-GAL4 and UAS-dsKu80^{JF02790} were used. The average times required to lose postural control due to ethanol intoxication were: *UAS-dsKu80^{JF02790}*/ + : 22.3 ± 1.7; *Appl-GAL4*/ + : 20.5 ± 1.8; and *UAS-dsKu80^{JF02790}*/Appl-GAL4: 27.8 ± 1.1. The number of experiments performed was 7, 7, and 8 respectively. Each experiment involved one fly population of $n = 100$ and the error bars represent SEM. (*) indicates significance $P < 0.05$ as determined with an ANOVA Tukey-Kramer *post hoc* test comparing experimental group with controls.
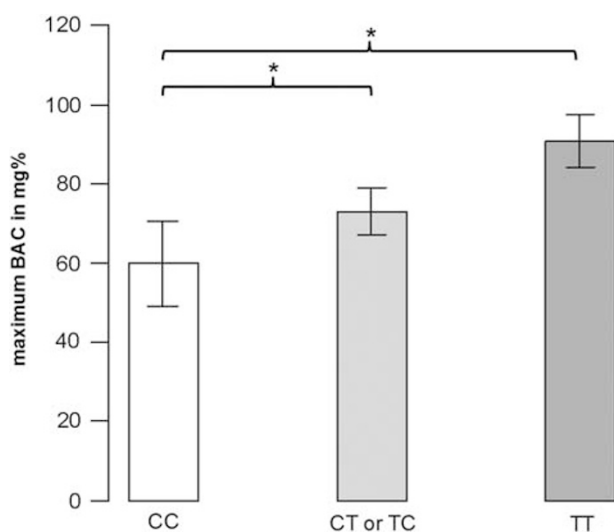
Figure 2 Effect of genotype on alcohol self-administration. Mean and standard error of the mean (SEM) of the highest blood alcohol concentration (BAC) achieved during 2 h of voluntary free access intravenous self-infusion of ethanol in CC (15), CT (43), and TT (27). The effect of genotype was statistically significant. Error bars represent standard errors of the mean. Significance code: *$P < 0.05$.

AD and alcohol-related traits, ie glycolysis/gluconeogenesis. According to KEGG, these processes include alcohol degradation and the conversion of the resulting acetate to acetyl CoA. In rat models, chronic alcohol consumption has been associated with changes in the enzymes of glycolysis and gluconeogenesis in the liver (Klouckova *et al*, 2006), as well as with enzymes of glycolysis in the hippocampus (Hargreaves *et al*, 2009). Research in humans has demonstrated inhibition of gluconeogenesis following alcohol ingestion, as determined through quantification of the gluconeogenic flux (Siler *et al*, 1998). The glycolysis/gluconeogenesis set contained two well-established groups of candidate genes, ie the alcohol dehydrogenases (*ADH1B* and *ADH1C* (for review see (Rietschel and Treutlein, 2013), see also www.addictiongwas.com described in (Spanagel *et al*, 2013)), and the aldehyde dehydrogenases (*ALDH1A3* (Sherva *et al*, 2009), *ALDH7A1* (Bhave *et al*, 2006). (Supplementary Table S1). The aldehyde dehydrogenases belong to the same gene family as *ALDH2*, which exerts strong protective effects against AD in Asian ethnicities (for review, see (Rietschel and Treutlein, 2013).

With respect to the identified gene-sets, glycolysis/gluconeogenesis was the only one to have received strong support from previous studies. We therefore sought independent support for the 38 most significant contributory genes ($P < 1E-3$) from our 19 identified gene-sets. In the gene-wide analysis of the COGA/SAGE replication data set, the genes *CLDN14* and *SLC2A13* achieved nominal significance. With respect to the non-replicated genes, it must be borne in mind that for complex traits, non-replication does not necessarily indicate that the initial results were false. This is underlined by the fact that in addition to *ADH1B* and *ADH1C*, 16 of the 38 genes in the top category of the 19 gene-sets have been reported as candidate genes for alcohol-related traits in previous studies, whereas 20 are new candidates (Table 2).

The fact that 15 of the 19 gene-sets achieved significance owing to variants that were not among the best single-marker findings ($P < 1E-5$) demonstrates the ability of the present approach to identify pathophysiologically relevant genes that are overlooked in single-marker analyses of GWAS data. This is also illustrated by the markers in *XRCC5*, all of which had *P*-values of $> 1E-5$.

The four top SNPs in *XRCC5* were only moderately correlated (rs828701-rs828704: $r^2 = 0.271$, D′ = 1.000; rs828701-rs207938: $r^2 = 0.257$, D′ = 0.533; rs828701-rs2032765: $r^2 = 0.016$, D′ = 0.424; rs828704-rs207938: $r^2 = 0.021$, D′ = 0.333; rs828704-rs2032765: $r^2 = 0.094$, D′ = 0.469; rs207938-rs2032765: $r^2 = 0.080$, D′ = 1.000; 1000 Genomes Pilot 1 CEU data accessed via https://www.broadinstitute.org/mpg/snap/ldsearch.php). Only the top SNP rs828701 was genotyped, as none of the four SNPs constituted an eQTL or an amino-acid exchange.

XRCC5 was present in approximately one third of the discovered gene-sets. The application of a convergent approach (Spanagel, 2013) generated strong further independent support for *XRCC5*. Previous gene expression studies in rodents have implicated *XRCC5* in the etiology of AD. Expression profiling in inbred long sleep/short sleep mouse strains—which display pronounced differences in ethanol sensitivity as measured by loss of the righting response (LORR)—identified 15 differentially expressed

Table 2 Previously Implicated Candidates Among the Thirty-Eight most Significant Contributory Genes (*P* ⩽ 1E-3) in the 19 Identified Gene-Sets

Gene	Location (Entrez gene cytogenetic band)	Gene product function	Gene-sets	References for previously implicated candidates
ADAMTS-like 1 (*ADAMTSL1*)	9p21	Extracellular matrix protein	18, 19	(Edenberg *et al*, 2010; Kendler *et al*, 2011)
Alcohol dehydrogenases 1B (*ADH1B*)	4q23	Alcohol degradation	9	(Li *et al*, 2011; McKay *et al*, 2011)
Alcohol dehydrogenase 1C (*ADH1C*)	4q23		9	(Li *et al*, 2012)
Aryl hydrocarbon receptor nuclear translocator-like (*ARNTL*)	11p15	Circadian rhythm	19	(Kovanen *et al*, 2010)
ATPase2A3 (*ATP2A3*)	17p13	Role in the hydrolysis of adenosine triphosphate/calcium transport	19	—
Lysosomal V0 subunit a4 of the H+ transporting ATPase (*ATP6V0A4*)	7q34	Involved in the acidification of intracellular compartments	10	—
Ataxin 1 (*ATXN1*)	6p23	Chromatin-binding factor that represses Notch signaling	13	—
Claudin 14 (*CLDN14*)	21q22	Component of tight junctions	8	—
Diacylglycerol kinase iota (*DGKI*)	7q32-q33	Participates in signal transduction by modulating levels of diacylglycerol	3, 17	(Carr *et al*, 2007; Liang *et al*, 2010; Sommer *et al*, 2001)
Dematin (*EPB49*)	8p21	Actin-bundling protein	13	—
FAT tumor suppressor homolog 2 (*FAT2*)	5q33	Encodes a cadherin with a role in the development of the cerebellum	13	—
F-box and leucine-rich repeat protein 17 (*FBXL17*)	5q21	Role in ubiquitination	15	(Kendler *et al*, 2011)
Fibroblast growth factor 17 (*FGF17*)	8p21	Role in cell growth and embryonic development	18	—
Forkhead box P1 (*FOXP1*)	3p14	Role in cell type- and tissue specific transcription	18, 19	—
Histone deacetylase 9 (*HDAC9*)	7p21	Deacetylatylase involved in epigenetic repression of transcription	13, 19	—
Homeodomain interacting protein kinase 2 (*HIPK2*)	7q32-q34	Serine/threonine-protein kinase involved in the regulation of transcription	17	(Mulligan *et al*, 2006)
Keratin 6B (*KRT6B*)	12q13	Role in the differentiation of epithelial tissues	14	—
Laminin alpha 2 (*LAMA2*)	6q22-q23	Regulates postnatal oligodendrogenesis	7	(Kendler *et al*, 2011; Mulligan *et al*, 2006)
Low-density lipoprotein receptor-related protein 1B (*LRP1B*)	2q21	Internalizes ligands during receptor-mediated endocytosis	19	(Edenberg *et al*, 2010; Johnson *et al*, 2006)
Leucine-rich repeat transmembrane neuronal 4 (*LRRTM4*)	2p12	Role in development of the nervous system	18	(Johnson *et al*, 2006)
Latent transforming growth factor beta binding protein 1 (*LTBP1*)	2p22-p21	Role in transforming growth factor beta signaling	19	(Pei *et al*, 2012)
Myosin IE (*MYO1E*)	15q21-q22	Molecular motor	18	(Mulligan *et al*, 2006)
Peroxisomal trans-2-enoyl-coenzyme A reductase (*PECR*)	2q35	May participate in fatty acid chain elongation	6, 12, 16	—
PKI55 gene (*PKI55*)	2q35	Protein kinase C modulator	16	—
Peptidylprolyl isomerase A (*PPIA*)	7p13	Accelerates the folding of polypeptides	2, 5	(Mulligan *et al*, 2006)
Protein kinase C alpha (*PRKCA*)	17q22-q23	Phosphorylates target proteins, involved in signaling	1, 7, 8	(Edenberg *et al*, 2010; Han *et al*, 2013; Mulligan *et al*, 2006; Rodd *et al*, 2008; Slater *et al*, 2003)
rap guanine nucleotide exchange factor 4 (*RAPGEF4*)	2q31-q32	Guanine nucleotide exchange factor for small guanosin triphosphatases	8	—
Reelin (*RELN*)	7q22	Role during the migration of neurons during brain development	7	(Kimpel *et al*, 2007; Rodd *et al*, 2008)
ras homolog family member H (*RHOH*)	4p13	Small guanosine triphosphatase involved in the regulation of cell proliferation and survival	3, 8	—
Ribosomal protein S20 pseudogene 3 (*RPS20P3*)	5q21	No known function	15	—
Solute carrier family 2 member 13 (*SLC2A13*)	12q12	Proton/myo-inositol cotransporter	14	—
Sulfatase modifying factor 1 (*SUMF1*)	3p26	Role in the hydrolysis of sulfate esters	10, 13	(Mulligan *et al*, 2006)
Sushi domain containing 1 (*SUSD1*)	9q31-q33	Protein of unknown function	18	—
Supervillin (*SVIL*)	10p11	Forms a link between the membrane and the actin cytoskeleton	18	—
TEA domain family member 1 (*TEAD1*)	11p15	Transcription factor involved in the control of organ size	13, 19	(Edenberg *et al*, 2010; Mulligan *et al*, 2006)
Transmembrane protein 169 (*TMEM169*)	2q35	Protein of unknown function	16	—
Transcriptional regulating factor 1 (*TRERF1*)	6p21-p12	Zinc-finger protein regulating transcription	19	—
X-ray repair complementing defective repair in Chinese hamster cells 5 (*XRCC5*)	2q35	Role in the repair of DNA double strand breaks	1, 2, 4, 5, 11, 16	(Mulligan *et al*, 2006)

genes, one of which was *XRCC5* (MacLaren *et al*, 2006). LORR refers to the inability of mice to roll back onto their abdomens after receiving a sedative dose of alcohol and being placed on their backs (Crabbe *et al*, 2010). LORR is the correlate of the level of response to alcohol in humans, which refers to the reactions of an individual following alcohol ingestion (eg feeling 'high', reduced motor coordination) (Joslyn *et al*, 2010). An early linkage study mapped low level of response to alcohol to several chromosomal regions. These included chr2q35 (Schuckit *et al*, 2001), which hosts the *XRCC5* gene. Furthermore, in a recent gene-set-enrichment study of the level of response, *XRCC5* was among the 173 loci that contributed to this phenotype (Joslyn *et al*, 2010).

However, recent research has shown that the role of alcohol sensitivity in the development of AD is more complex (Newlin and Renton, 2010), and that responses to alcohol, such as greater pleasurable and excitatory effects, can increase the risk of future alcohol problems (King *et al*, 2011; King *et al*, 2014).

Therefore the issues of whether, and under which circumstances, level of response predicts AD remain unclear. However, the result of the present *Drosophila* study corroborates previous observations that *XRCC5* modifies the acute response to alcohol.

The present human genetic study provides further evidence for an association between *XRCC5* and effects of alcohol consumption in healthy young adult social drinkers.

The *XRCC5* variant rs828701, which was the most significant *XRCC5* finding in our GWAS of AD ($P = 2.25E-5$), showed an allele-dosage-dependent association with the maximum achieved BAC in an experiment, in which volunteers used a free access alcohol self-infusion paradigm to reproduce the pleasant alcohol effects preferred within the context of a weekend party.

Compared with the findings described above, the results of this experiment are not only influenced by the level of response to inebriating effects of alcohol, but also by the behaviorally relevant constructs of 'liking' and 'wanting' alcohol, according to the theory of incentive sensitization (Robinson and Berridge, 2000). However, association was found with the T allele, and not with the AD risk allele C.

Rs828701 does not constitute an amino-acid exchange, and has no known effect on gene expression. Thus the possibility that this represents a chance finding cannot be excluded. However, we consider the finding genuine, and attribute it to the well-known phenomenon where by different alleles of the same variant are associated with the same phenotype. This phenomenon was first described in 2007 (Lin *et al*, 2007), and has been observed for some of the most robust genetic association findings in both humans (Lin *et al*, 2007; Maher *et al*, 2010), and *Drosophila* (Gruber *et al*, 2007). These authors proposed that this phenomenon may be explained by an interaction or correlation between the examined (proxy) variant and an unknown causal variant. Events of this nature are most likely to occur when proxies are examined in populations with different genetic backgrounds. All of the present study participants were of German descent. Nevertheless, certain differences between cohorts may have introduced some unrecognized difference in genetic composition: The risk allele was found in AD patients (Frank *et al*, 2012) with an average age of 42.0 years who had commenced drinking many years ago, whereas the study of alcohol self-administration was conducted in healthy, non-alcohol-dependent younger social drinkers with an average age of 18.39 years. Furthermore, research has shown that light social drinkers could be a unique group, as their drinking over follow-up remained light, indicating the presence of factors that mitigate their risk for heavy drinking (King *et al*, 2014).

The effect may have also been due to a gene x environment (GxE) interaction. The presence of a GxE means that the effect of a gene differs between different environments. This may also explain inconsistent findings in the literature concerning whether increased alcohol sensitivity confers an increased risk of AD. Unrecognized, yet important, environmental factors may act in our study populations. More detailed comparison of the alcohol intake characteristics of controls and the alcohol self-administration group was hampered by the fact that the controls were population based, as are many GWAS control samples. The possibility that the reversed allele effect was due to the chance occurrence of erratic allele frequencies can be excluded by the finding that the allele frequency in the non-alcohol-dependent social drinkers (C-allele frequency 44.9%) was closer to that of the controls (C-allele frequencies: 43.5%) than to that of the cases (C-allele frequency: 48.6%).

XRCC5 encodes a protein involved in the repair of DNA double strand breaks (Downs and Jackson, 2004). *XRCC5* homologs are involved in similar phenotypes in the organismal lineages human, mouse, and fly, whose last common ancestor existed around 970 million years ago (Nei *et al*, 2001). This suggests that the function of *XRCC5/Ku80* in the cellular context, which is presumably related to double strand break repair, has been preserved owing to its vital importance. *XRCC5*'s involvement in neuronal homeostasis (De Zio *et al*, 2012) may also influence the manner in which an organism reacts to ethanol.

A potential limitation of our gene-set-based study is that although correction was made for gene-set significance within each gene-set database, no correction was made for all databases in total. This increases the likelihood of false-positive findings. The procedure was chosen to limit the risk of rejecting true findings on the basis of their small effect sizes and indeed we could replicate 5 out of the 19 gene-sets in an independent sample. However, the other results require replication in future independent studies.

A further issue concerns comparability between alcohol sensitivity in model organisms and response to alcohol in humans (Crabbe *et al*, 2010). Although the comparability of these phenotypes is open to question, human *XRCC5* and *D. melanogaster Ku80* show 1:1 orthology, and thus comparability in terms of the corresponding gene is strong.

Despite uncertainty concerning phenotypic correlation between species, preliminary analyses of the biological mechanisms that are associated with alcohol-induced behaviors in both species (eg cAMP signaling) suggest that the ethanol phenotypes in *Drosophila* may indeed serve as a proxy for more complex alcohol-induced behaviors in humans (Moore *et al*, 1998, Schuckit *et al*, 2004). In particular, the correlation between alcohol sensitivity/resistance in *Drosophila* and the level of response in humans is supported by convincing molecular data (Lasek *et al*, 2011).

In conclusion, the results of the present gene-set-based analysis identified known and new candidate genes for AD. Our follow-up study of *XRCC5* in *Drosophila* and human endorsed previous convergent findings from gene expression studies in rodents, and findings from linkage- and gene-set-enrichment studies in humans, in suggesting that this gene impacts on response to alcohol. Further studies are warranted to replicate these candidate genes, and to elucidate their precise mechanisms of action and their relevance in AD.

REFERENCES

Altenhoff AM, Gil M, Gonnet GH, Dessimoz C (2013). Inferring hierarchical orthologous groups from orthologous gene pairs. *PLoS One* 8: e53786.

Baik I, Cho NH, Kim SH, Han BG, Shin C (2011). Genome-wide association studies identify genetic loci related to alcohol consumption in Korean men. *Am J Clin Nutr* 93: 809–816.

Bierut LJ, Agrawal A, Bucholz KK, Doheny KF, Laurie C, Pugh E et al (2010). A genome-wide association study of alcohol dependence. *Proc Natl Acad Sci USA* 107: 5082–5087.

Bhave SV, Hoffman PL, Lassen N, Vasiliou V, Saba L, Deitrich RA et al (2006). Gene array profiles of alcohol and aldehyde metabolizing enzymes in brains of C57BL/6 and DBA/2 mice. *Alcohol Clin Exp Res* 30: 1659–1669.

Carr LG, Kimpel MW, Liang T, McClintick JN, McCall K, Morse M et al (2007). Identification of candidate genes for alcohol preference by expression profiling of congenic rat strains. *Alcohol Clin Exp Res* 31: 1089–1098.

Crabbe JC, Bell RL, Ehlers CL (2010). Human and laboratory rodent low response to alcohol: is better consilience possible? *Addict Biol* 15: 125–144.

De Zio D, Bordi M, Cecconi F (2012). Oxidative DNA damage in neurons: implication of ku in neuronal homeostasis and survival. *Int J Cell Biol* 2012: 752420.

Deelen J, Uh HW, Monajemi R, van Heemst D, Thijssen PE, Bohringer S et al (2013). Gene set analysis of GWAS data for human longevity highlights the relevance of the insulin/IGF-1 signaling and telomere maintenance pathways. *Age (Dordr)* 35: 235–249.

Downs JA, Jackson SP (2004). A means to a DNA end: the many roles of Ku. *Nat Rev Mol Cell Biol* 5: 367–378.

Edenberg HJ, Koller DL, Xuei X, Wetherill L, McClintick JN, Almasy L et al (2010). Genome-wide association study of alcohol dependence implicates a region on chromosome 11. *Alcohol Clin Exp Res* 34: 840–852.

Frank J, Cichon S, Treutlein J, Ridinger M, Mattheisen M, Hoffmann P et al (2012). Genome-wide significant association between alcohol dependence and a variant in the *ADH* gene cluster. *Addict Biol* 17: 171–180.

Gelernter J, Kranzler HR, Sherva R, Almasy L, Koesterer R, Smith AH et al (2013a). Genome-wide association study of alcohol dependence: significant findings in African- and European-Americans including novel risk loci. *Mol Psychiatry* 19: 41–49.

Gelernter J, Kranzler HR, Sherva R, Koesterer R, Almasy L, Zhao H et al (2013b). Genome-wide association study of opioid dependence: multiple associations mapped to calcium and potassium pathways. *Biol Psychiatry* 76: 66–74.

Gelernter J, Sherva R, Koesterer R, Almasy L, Zhao H, Kranzler HR et al (2013c). Genome-wide association study of cocaine dependence and related traits: *FAM53B* identified as a risk gene. *Mol Psychiatry* 19: 717–723.

Convergent genetic studies implicate *XRCC5* in AD
D Juraeva *et al*

370

Goeman JJ, van de Geer SA, de Kort F, van Houwelingen HC (2004). A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics* 20: 93–99.

Gruber JD, Genissel A, Macdonald SJ, Long AD (2007). How repeatable are associations between polymorphisms in achaete-scute and bristle number variation in *Drosophila*? *Genetics* 175: 1987–1997.

Han S, Yang BZ, Kranzler HR, Liu X, Zhao H, Farrer LA *et al* (2013). Integrating GWASs and human protein interaction networks identifies a gene subnetwork underlying alcohol dependence. *Am J Hum Genet* 93: 1027–1034.

Hargreaves GA, Quinn H, Kashem MA, Matsumoto I, McGregor IS (2009). Proteomic analysis demonstrates adolescent vulnerability to lasting hippocampal changes following chronic alcohol consumption. *Alcohol Clin Exp Res* 33: 86–94.

Hoffmann AA, Cohan FM (1987). Olfactory responses of *Drosophila melanogaster* selected for knockdown resistance to ethanol. *Behav Genet* 17: 307–312.

Jia P, Zheng S, Long J, Zheng W, Zhao Z (2011). dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks. *Bioinformatics* 27: 95–102.

Johnson C, Drgon T, Liu QR, Walther D, Edenberg H, Rice J *et al* (2006). 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* 141B: 844–853.

Joslyn G, Ravindranathan A, Brush G, Schuckit M, White RL (2010). Human variation in alcohol response is influenced by variation in neuronal signaling genes. *Alcohol Clin Exp Res* 34: 800–812.

Juraeva D, Haenisch B, Zapatka M, Frank J, GROUP Investigators, iPSYCH-GEMS SCZ working group *et al* (2014). Integrated pathway-based approach identifies association between genomic regions at *CTCF* and *CACNB2* and schizophrenia. *PLoS Genet* 10: e1004345.

Kendler KS, Kalsi G, Holmans PA, Sanders AR, Aggen SH, Dick DM *et al* (2011). Genomewide association analysis of symptoms of alcohol dependence in the molecular genetics of schizophrenia (MGS2) control sample. *Alcohol Clin Exp Res* 35: 963–975.

Kimpel MW, Strother WN, McClintick JN, Carr LG, Liang T, Edenberg HJ *et al* (2007). Functional gene expression differences between inbred alcohol-preferring and -non-preferring rats in five brain regions. *Alcohol* 41: 95–132.

King AC, de Wit H, McNamara PJ, Cao D (2011). Rewarding, stimulant, and sedative alcohol responses and relationship to future binge drinking. *Arch Gen Psychiatry.* 68: 389–399.

King AC, McNamara PJ, Hasin DS, Cao D (2014). Alcohol challenge responses predict future alcohol use disorder symptoms: a 6-year prospective study. *Biol Psychiatry.* 75: 798–806.

Klouckova I, Hrncirova P, Mechref Y, Arnold RJ, Li TK, McBride WJ *et al* (2006). Changes in liver protein abundance in inbred alcohol-preferring rats due to chronic alcohol exposure, as measured through a proteomics approach. *Proteomics* 6: 3060–3074.

Kovanen L, Saarikoski ST, Haukka J, Pirkola S, Aromaa A, Lonnqvist J *et al* (2010). Circadian clock gene polymorphisms in alcohol use disorders and alcohol consumption. *Alcohol Alcohol* 45: 303–311.

Lasek AW, Lim J, Kliethermes CL, Berger KH, Joslyn G, Brush G *et al* (2011). An evolutionary conserved role for anaplastic lymphoma kinase in behavioral responses to ethanol. *PLoS One* 6: e22636.

Li D, Zhao H, Gelernter J (2011). Strong association of the alcohol dehydrogenase 1B gene (*ADH1B*) with alcohol dependence and alcohol-induced medical diseases. *Biol Psychiatry* 70: 504–512.

Li D, Zhao H, Gelernter J (2012). Further clarification of the contribution of the *ADH1C* gene to vulnerability of alcoholism and selected liver diseases. *Hum Genet* 131: 1361–1374.

Li H, Coghlan A, Ruan J, Coin LJ, Heriche JK, Osmotherly L *et al* (2006). TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res* 34(Database issue): D572–D580.

Liang T, Kimpel MW, McClintick JN, Skillman AR, McCall K, Edenberg HJ *et al* (2010). Candidate genes for alcohol preference identified by expression profiling in alcohol-preferring and -nonpreferring reciprocal congenic rats. *Genome Biol* 11: R11.

Lin PI, Vance JM, Pericak-Vance MA, Martin ER (2007). No gene is an island: the flip-flop phenomenon. *Am J Hum Genet* 80: 531–538.

MacLaren EJ, Bennett B, Johnson TE, Sikela JM (2006). Expression profiling identifies novel candidate genes for ethanol sensitivity QTLs. *Mamm Genome* 17: 147–156.

Maher BS, Reimers MA, Riley BP, Kendler KS (2010). Allelic heterogeneity in genetic association meta-analysis: an application to *DTNBP1* and schizophrenia. *Hum Hered* 69: 71–79.

McKay JD, Truong T, Gaborieau V, Chabrier A, Chuang SC, Byrnes G *et al* (2011). A genome-wide association study of upper aerodigestive tract cancers conducted within the INHANCE consortium. *PLoS Genet* 7: e1001333.

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

Mulligan MK, Ponomarev I, Hitzemann RJ, Belknap JK, Tabakoff B, Harris RA *et al* (2006). Toward understanding the genetics of alcohol drinking through transcriptome meta-analysis. *Proc Natl Acad Sci USA* 103: 6368–6373.

Nei M, Xu P, Glazko G (2001). Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. *Proc Natl Acad Sci USA* 98: 2497–2502.

Newlin DB, Renton RM (2010). High risk groups often have higher levels of alcohol response than low risk: the other side of the coin. *Alcohol Clin Exp Res* 34: 199–202.

Pan Y, Luo X, Liu X, Wu LY, Zhang Q, Wang L *et al* (2013). Genome-wide association studies of maximum number of drinks. *J Psychiatr Res* 47: 1717–1724.

Park BL, Kim JW, Cheong HS, Kim LH, Lee BC, Seo CH *et al* (2013). Extended genetic effects of *ADH* cluster genes on the risk of alcohol dependence: from GWAS to replication. *Hum Genet* 132: 657–668.

Pei YF, Zhang L, Yang TL, Han Y, Hai R, Ran S *et al* (2012). Genome-wide association study of copy number variants suggests *LTBP1* and *FGD4* are important for alcohol drinking. *PLoS One* 7: e30860.

Prescott CA, Kendler KS (1999). Genetic and environmental contributions to alcohol abuse and dependence in a population-based sample of male twins. *Am J Psychiatry* 156: 34–40.

Quillen EE, Chen XD, Almasy L, Yang F, He H, Li X *et al* (2014). ALDH2 is associated to alcohol dependence and is the major genetic determinant of 'daily maximum drinks' in a GWAS study of an isolated rural Chinese sample. *Am J Med Genet B Neuropsychiatr Genet* 165B: 103–110.

Rietschel M, Treutlein J (2013). The genetics of alcohol dependence. *Ann NY Acad Sci* 1282: 39–70.

Robinson TE, Berridge KC (2000). The psychology and neurobiology of addiction: an incentive-sensitization view. *Addiction* 95(Suppl 2): S91–117.

Rodd ZA, Kimpel MW, Edenberg HJ, Bell RL, Strother WN, McClintick JN *et al* (2008). Differential gene expression in the nucleus accumbens with ethanol self-administration in inbred alcohol-preferring rats. *Pharmacol Biochem Behav* 89: 481–498.

Scholz H (2009). Intoxicated fly brains: neurons mediating ethanol-induced behaviors. *J Neurogenet* 23: 111–119.

Schuckit MA (1994). Low level of response to alcohol as a predictor of future alcoholism. *Am J Psychiatry* 151: 184–189.

Schuckit MA, Edenberg HJ, Kalmijn J, Flury L, Smith TL, Reich T et al (2001). A genome-wide search for genes that relate to a low level of response to alcohol. *Alcohol Clin Exp Res* 25: 323–329.

Schuckit MA, Smith TL, Kalmijn J (2004). The search for genes contributing to the low level of response to alcohol: patterns of findings across studies. *Alcohol Clin Exp Res* 28: 1449–1458.

Schumann G, Coin LJ, Lourdusamy A, Charoen P, Berger KH, Stacey D et al (2011). Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (*AUTS2*) in the regulation of alcohol consumption. *Proc Natl Acad Sci USA* 108: 7119–7124.

Sherva R, Rice JP, Neuman RJ, Rochberg N, Saccone NL, Bierut LJ (2009). Associations and interactions between SNPs in the alcohol metabolizing genes and alcoholism phenotypes in European Americans. *Alcohol Clin Exp Res* 33: 848–857.

Siler SQ, Neese RA, Christiansen MP, Hellerstein MK (1998). The inhibition of gluconeogenesis following alcohol in humans. *Am J Physiol* 275(5 Pt 1): E897–E907.

Slater SJ, Cook AC, Seiz JL, Malinowski SA, Stagliano BA, Stubbs CD (2003). Effects of ethanol on protein kinase C alpha activity induced by association with Rho GTPases. *Biochemistry* 42: 12105–12114.

Sommer W, Arlinde C, Caberlotto L, Thorsell A, Hyytia P, Heilig M (2001). Differential expression of diacylglycerol kinase iota and L18A mRNAs in the brains of alcohol-preferring AA and alcohol-avoiding ANA rats. *Mol Psychiatry* 6: 103–108 105.

Spanagel R, Durstewitz D, Hansson A, Heinz A, Kiefer F, Kohr G et al (2013). A systems medicine research approach for studying alcohol addiction. *Addict Biol* 18: 883–896.

Spanagel R (2013). Convergent functional genomics in addiction research—a translational approach to study candidate genes and gene networks. *In Silico Pharmacology* 1: 18.

Sullivan PF, Daly MJ, O'Donovan M (2012). Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat Rev Genet* 13: 537–551.

Takeuchi F, Isono M, Nabika T, Katsuya T, Sugiyama T, Yamaguchi S et al (2011). Confirmation of *ALDH2* as a Major locus of drinking behavior and of its variants regulating multiple metabolic phenotypes in a Japanese population. *Circ J* 75: 911–918.

Treutlein J, Cichon S, Ridinger M, Wodarz N, Soyka M, Zill P et al (2009). Genome-wide association study of alcohol dependence. *Arch Gen Psychiatry* 66: 773–784.

Veyrieras JB, Kudaravalli S, Kim SY, Dermitzakis ET, Gilad Y, Stephens M et al (2008). High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet* 4: e1000214.

Wetherill L, Kapoor M, Agrawal A, Bucholz K, Koller D, Bertelsen SE et al (2013). Family-based association analysis of alcohol dependence criteria and severity. *Alcohol Clin Exp Res* 38: 354–366.

Zimmermann US, Mick I, Laucht M, Vitvitskiy V, Plawecki MH, Mann KF et al (2009). Offspring of parents with an alcohol use disorder prefer higher levels of brain alcohol exposure in experiments involving computer-assisted self-infusion of ethanol (CASE). *Psychopharmacology* 202: 689–697.

Zuo L, Wang K, Zhang XY, Krystal JH, Li CS, Zhang F et al (2013). *NKAIN1-SERINC2* is a functional, replicable and genome-wide significant risk gene region specific for alcohol dependence in subjects of European descent. *Drug Alcohol Depend* 129: 254–264.

Supplementary Information accompanies the paper on the Neuropsychopharmacology website (http://www.nature.com/npp)