

ORIGINAL ARTICLE

Genome-wide association analysis of coffee drinking suggests association with *CYP1A1/CYP1A2* and *NRCAM*

N Amin^{1,38}, E Byrne^{2,38}, J Johnson², G Chenevix-Trench², S Walter^{3,4}, IM Nolte⁵, kConFab Investigators⁶, JM Vink⁷, R Rawal⁸, M Mangino⁹, A Teumer¹⁰, JC Keers¹¹, G Verwoert³, S Baumeister¹², R Biffar¹³, A Petersmann¹⁴, N Dahmen¹⁵, A Doering⁸, A Isaacs¹, L Broer¹, NR Wray², GW Montgomery², D Levy^{16,17}, BM Psaty^{18,19}, V Gudnason^{20,21}, A Chakravarti^{22,23}, P Sulem²⁴, DF Gudbjartsson²⁴, LA Kiemeneij^{25,26,27}, U Thorsteinsdottir^{24,28}, K Stefansson^{24,28}, FJA van Rooij⁴, YS Aulchenko¹, JJ Hottenga⁷, FR Rivadeneira²⁹, A Hofman³, AG Uitterlinden²⁹, CJ Hammond³⁰, S-Y Shin³⁰, A Ikram³, JCM Witteman³, ACJW Janssens³, H Snieder^{5,11}, H Tiemeier^{3,31}, BHR Wolfenbuttel^{11,32}, BA Oostra^{1,33}, AC Heath³⁴, E Wichmann^{8,35}, TD Spector⁹, HJ Grabe³⁶, DI Boomsma⁷, NG Martin^{2,38} and CM van Duijn^{1,37,38}

¹Unit of Genetic Epidemiology, Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands; ²Department of Genetics, Queensland Institute of Medical Research, Brisbane, Queensland, Australia; ³Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands; ⁴Department of Public Health, Erasmus University Medical Center, Rotterdam, The Netherlands; ⁵Unit of Genetic Epidemiology and Bioinformatics, Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; ⁶Peter MacCallum Cancer Institute, Melbourne, Victoria, Australia; ⁷Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands; ⁸Institute of Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany; ⁹Department of Twin Research and Genetic Epidemiology, St Thomas' Hospital Campus, King's College London, London, UK; ¹⁰Interfaculty Institute for Genetics and Functional Genomics, Department of Functional Genomics, University of Greifswald, Greifswald, Germany; ¹¹LifeLines Cohort Study and Biobank, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; ¹²Institute for Community Medicine, University of Greifswald, Greifswald, Germany; ¹³Department of Prosthodontics, Gerodontology and Dental Materials, Center of Oral Health, University of Greifswald, Greifswald, Germany; ¹⁴Institute of Clinical Chemistry and Laboratory Medicine, University of Greifswald, Greifswald, Germany; ¹⁵Department of Psychiatry, University of Mainz, Mainz, Germany; ¹⁶National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA, USA; ¹⁷Center for Population Studies, NHLBI, Bethesda, MD, USA; ¹⁸Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA, USA; ¹⁹Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA; ²⁰Icelandic Heart Association, Kopavogur, Iceland; ²¹University of Iceland, Reykjavik, Iceland; ²²McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD, USA; ²³Department of Epidemiology and Medicine, Johns Hopkins University, Baltimore, MD, USA; ²⁴deCODE Genetics, Reykjavik, Iceland; ²⁵Department of Urology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; ²⁶Department of Endocrinology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands; ²⁷Comprehensive Cancer Center East, BG Nijmegen, The Netherlands; ²⁸Faculty of Medicine, University of Iceland, Reykjavik, Iceland; ²⁹Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands; ³⁰Human Genetics, Wellcome Trust Sanger Institute, Genome Campus, Hinxton, UK; ³¹Department of Child and Adolescent Psychiatry, Erasmus University Medical Center, Rotterdam, The Netherlands; ³²Department of Endocrinology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; ³³Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands; ³⁴Department of Psychiatry, Washington University, St Louis, MI, USA; ³⁵Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany; ³⁶Department of Psychiatry and Psychotherapy, University of Greifswald, Stralsund, Germany and ³⁷Centre of Medical Systems Biology, Netherlands Consortium on Healthy Aging, Leiden and National Genomics Initiative, The Hague, The Netherlands

Coffee consumption is a model for addictive behavior. We performed a meta-analysis of genome-wide association studies (GWASs) on coffee intake from 8 Caucasian cohorts (N=18176) and sought replication of our top findings in a further 7929 individuals. We also

Correspondence: Professor CM van Duijn, Department of Epidemiology, Erasmus Medical Center Rotterdam, Dr Molewaterplein 50, 3015 GE, Rotterdam, The Netherlands.

E-mail: c.vanduijn@erasmusmc.nl

³⁸These authors contributed equally to this work.

Received 2 February 2011; revised 28 June 2011; accepted 7 July 2011; published online 30 August 2011

performed a gene expression analysis treating different cell lines with caffeine. Genome-wide significant association was observed for two single-nucleotide polymorphisms (SNPs) in the 15q24 region. The two SNPs rs2470893 and rs2472297 (P -values = 1.6×10^{-11} and 2.7×10^{-11}), which were also in strong linkage disequilibrium ($r^2 = 0.7$) with each other, lie in the 23-kb long commonly shared 5' flanking region between *CYP1A1* and *CYP1A2* genes. *CYP1A1* was found to be downregulated in lymphoblastoid cell lines treated with caffeine. *CYP1A1* is known to metabolize polycyclic aromatic hydrocarbons, which are important constituents of coffee, whereas *CYP1A2* is involved in the primary metabolism of caffeine. Significant evidence of association was also detected at rs382140 (P -value = 3.9×10^{-09}) near *NRCAM*—a gene implicated in vulnerability to addiction, and at another independent hit rs6495122 (P -value = 7.1×10^{-09})—an SNP associated with blood pressure—in the 15q24 region near the gene *ULK3*, in the meta-analysis of discovery and replication cohorts. Our results from GWASs and expression analysis also strongly implicate *CAB39L* in coffee drinking. Pathway analysis of differentially expressed genes revealed significantly enriched ubiquitin proteasome (P -value = 2.2×10^{-05}) and Parkinson's disease pathways (P -value = 3.6×10^{-05}).

Molecular Psychiatry (2012) 17, 1116–1129; doi:10.1038/mp.2011.101; published online 30 August 2011

Keywords: coffee; P450; NRCAM; CAB39L; Parkinson's disease; CYP1A1/CYP1A2

Introduction

Coffee is the most widely used beverage worldwide with known health benefits.^{1–3} Coffee drinking has been associated with a decreased risk of dementia and Alzheimer's disease (AD), Parkinson's disease⁴ and type II diabetes.⁵ However, coffee intake has also been associated with increased risk of some cancers,^{6–8} blood pressure^{9–10} and myocardial infarction.¹¹ The contradictory epidemiological findings on coffee intake and its effects on various diseases may be attributed to the different contents of coffee. Apart from caffeine, its most well-known constituent that stimulates the central nervous system, coffee is a source of complex organic compounds with beneficial anti-oxidant and endocrine properties.^{12,13} However, roasting of coffee beans is known to produce polycyclic aromatic hydrocarbons,¹⁴ which are a class of carcinogenic chemicals that are formed by incomplete combustion of organic matter. At unusually high doses, coffee is known to have potentiating effects on mutagenesis,¹³ including cytotoxicity of X-rays, ultraviolet light and chemotherapeutic agents.¹³ Most of these mutagenic effects are known to be independent of caffeine¹⁵ and have been attributed to aliphatic dicarbonyls^{15–17} and hydrogen peroxide.¹⁶

The P450 system in the liver has a key role in coffee metabolism. The cytochrome P450A1 encoded by the gene *CYP1A1* is known to metabolize polycyclic aromatic hydrocarbons such as benzo(a)pyrene. The caffeine content (~100 mg per cup of coffee) is primarily metabolized in the liver by the cytochrome P450 *CYP1A2*, and is further broken down by the enzymes *CYP2A6* and *NAT2*.^{18–21} Most of the biological effects of caffeine including those on the brain and the central nervous system are mediated through antagonism of the adenosine receptors, specifically the A1 and A2A receptors.²² Accumulating evidence from a number of sources points to the A2A receptor as the main target for caffeine.²³ Under normal conditions, adenosine is hypothesized to activate the adenosine receptors, leading to

subsequent activation of adenylyl cyclase and Ca^{2+} channels.²² Adenosine acts to inhibit the release of neurotransmitters. Antagonism of the adenosine pathway leads to many downstream changes including changes in the dopaminergic system that result in decreased affinity of dopamine for the dopamine receptors²⁴ and changes in gene expression. It is through these mechanisms that caffeine mediates its effects on the brain and behavior.²² These changes in gene expression are not well understood, but the use of microarrays may give insights into the molecular changes brought about by caffeine.²⁵

Ordinary caffeine use has generally not been considered to be a case of drug abuse, and is indeed not so classified in DSM-IV (Diagnostic and Statistical Manual of Mental Disorder), but caffeine may be a model drug for studies of abuse,²⁶ and withdrawal effects when coffee consumption is stopped have been discussed. Many users experience withdrawal symptoms, which include headache, decreased alertness and concentration, as well as depressed mood and irritability.^{27,28} Beneficial effects from caffeine include improved psychomotor speed,²⁹ mood³⁰ and alertness.³¹ Several studies have shown that subjects reported higher levels of alertness and concentration along with increased appetite for work.^{22,30} Some have postulated that the positive effects can be explained by a reversal of the withdrawal symptoms in habitual users,^{32,33} but this has been refuted by studies of caffeine intake in non-habitual users.^{34,35} It is also clear that some users experience negative effects such as insomnia, anxiety and dysphoria.^{36,37} Caffeine consumption is also known to have a role in suicide.³⁸

Genetic studies in twins suggest that the heritability estimates of coffee consumption range from 0.39 to 0.56.^{30,39,40} Most genetic studies have focused on caffeine and have restricted the gene search primarily to polymorphisms in the *CYP1A2* and *ADORA2A* genes. Metabolism of caffeine by the *CYP1A2* enzyme shows substantial variation between people,^{41,42} because of both genetic and environmental factors.⁴³ There is some evidence, although not genome-wide

significant, that polymorphisms in the gene are known to moderate the association between coffee consumption and hypertension⁴⁴ and myocardial infarction,¹¹ as well as the risk of breast cancer in BRCA1 carriers.⁴⁵ No association has been found between variants in *CYP1A2* and caffeine consumption,²³ but a single-nucleotide polymorphism (SNP) in this gene (rs762551) has been shown to be associated with high inducibility of the *CYP1A2* enzyme in smokers.⁴⁶ A candidate gene study associated a polymorphism in the *ADORA2A* gene (rs5751876) with caffeine consumption²³ in a Costa Rican sample. This SNP has also been implicated in increasing risk to panic disorder in two separate studies in Caucasian populations,^{47,48} but none of the findings on *ADORA2A* reached genome-wide significance nor were they replicated. Among environmental factors, age and sex are known to affect coffee consumption.²⁸ Coffee consumption patterns may also differ from country to country based on its geographical location, religious preferences, availability of coffee and its cost; however, large differences remain between individuals within populations, presumably reflecting coffee preference, which is also known to be influenced by genetic variation.²⁸

We conducted a meta-analysis of genome-wide association studies (GWASs) including >18 000 individuals to identify common genetic variants that influence coffee consumption. We tested for association with >2.6 million polymorphisms that tag the vast majority of common human genetic variation. We combined the GWAS results with gene expression data from cells differentially treated with caffeine to identify genes the pattern of expression of which is changed after caffeine treatment and which harbor polymorphisms that show evidence of association.

Materials and methods

Study populations

This study included participants from eight cohorts including the following:

Erasmus Rucphen Family. The ERF (Erasmus Rucphen Family) study is a family-based study that includes >3000 participants descending from 22 couples living in the Rucphen region in the nineteenth century. All living descendants of these couples and their spouses were invited to participate in the study. The frequency of coffee consumption was assessed with a questionnaire. A total of 1814 participants who had both phenotype and genome-wide genotype data (54% women)⁴⁹ were available for the analysis.

Cooperative health research in the Augsburg region (KORA). The KORA F4 study is a follow-up study to the KORA-Survey 2000 (S4, October 1999 to July 2001). It was conducted between October 2006 and May 2008. From the KORA F4 survey (full cohort $n=3080$), 1814 individuals aged between 32 and 81

years were selected for genotyping on the Affymetrix 1000K⁵⁰ chip (Affymetrix, Santa Clara, CA, USA). Coffee consumption was assessed with a questionnaire asking the number of cups of coffee consumed per day.

Rotterdam Study (RS-I and RS-II). The Rotterdam Study-I (RS-I) is a prospective population-based cohort study of 7983 residents aged ≥ 55 years living in Ommoord, a suburb of Rotterdam (The Netherlands). Coffee consumption was assessed with a food frequency questionnaire. In total, 4139 individuals who had both phenotype and genotype data were used in the analysis.⁵¹ The RS-II is a prospective population-based cohort study of 3011 residents aged ≥ 55 years, and coffee consumption was assessed in the same manner.

Netherlands Twin Register. A sample of (mostly) adult twins was obtained from the NTR (Netherlands Twin Register), which was established in 1987 and contains information about Dutch twins and their families voluntarily participating in research.⁵² Since 1991, every 2–3 years, a questionnaire is mailed to adult twins and their family members registered with the NTR. These questionnaires contain items about health, lifestyle and personality. In 2000, the fifth NTR survey was sent out,⁵³ and contained the question, ‘On average, how many cups of caffeinated coffee do you drink in one day?’ This survey was completed by 6782 subjects; data on coffee consumption were available for 6673 subjects. The mean age of the respondents was 30.0 years (s.d. 10.9). Genome-wide genotyping was available for 1087 subjects with coffee data.

The Study of Health in Pomerania. The SHIP (Study of Health in Pomerania) is a longitudinal population-based cohort study in West Pomerania. The baseline sample SHIP-0 comprised 4308 subjects.^{54,55} Coffee consumption was assessed with the question, ‘How many cups of caffeinated coffee do you drink per day?’ In total, 2125 individuals (77.4% women) with both phenotype and genome-wide genotype data were available for the analysis.

TwinsUK. The TwinsUK cohort consisted of a group of twins ascertained to study the heritability and genetics of age-related diseases (<http://www.twinsUK.ac.uk>). These unselected twins were recruited from the general population through national media campaigns in the United Kingdom and were shown to be comparable to age-matched population singletons in terms of disease-related and lifestyle characteristics.^{56,57} The TwinsUK I and II cohorts consist of twins from the adult twin British registry, also shown to be representative of singleton populations and UK population.⁵⁶ Coffee consumption was assessed by questionnaire, and genome-wide genotype and imputed data were available for 1092 (TwinsUKI) and 1919 (TwinsUKII) samples. Ethics approval was obtained from the Guy’s

and St Thomas' Hospital Ethics Committee. Written informed consent was obtained from every participant in the study.

Queensland Institute of Medical Research (QIMR). Twins recruited from the Australian Twin Registry were mailed to a Health and Lifestyle Questionnaire between 1980 and 1982. Twins were recruited through national media campaigns. As part of the questionnaire, participants were asked 'How many cups of coffee would you drink on average per day?'²⁸ The age range of respondents was 17–88 years with a mean age of 31 years. Both phenotype and genotype information were available for 1988 unrelated individuals. Detailed descriptions of phenotyping, genotyping, imputation and quality control (QC) protocol are given elsewhere.^{28,58}

Genotyping and imputation

The participating cohorts were genotyped on commercially available platforms including Affymetrix, Illumina (San Diego, CA, USA) and Perlegen (Mountain View, CA, USA) (Supplementary Table 1). Quality control was performed in each group separately. The overall criteria were to exclude individuals with low call rate, excess heterozygosity and gender mismatch. On the basis of sample size and study-specific characteristics, different criteria were used (Supplementary Table 1). Imputations of non-genotyped SNPs in the HapMap CEU v22 were carried out within each study using MACH^{59,60} or IMPUTE.^{61,62}

Genome-wide association and meta analysis

We conducted a meta-analysis of 10 GWASs from 8 cohorts, consisting of >18 000 individuals and ~2.6 million imputed and genotyped SNPs. For each GWAS, the association analysis was performed using linear regression analysis by regressing coffee categories⁶³ on age, sex and SNP allele dosage in ProbABEL⁶⁴ (GenABEL.org, Novosibirsk, Russia), SNPTEST⁶⁵ (Jonathan Marchini, Oxford, UK) or QUICKTEST (Toby Johnson, Lausanne, Switzerland) (<http://toby.freeshell.org/software/quicktest.shtml>) (Supplementary Table 1). The coffee categories were defined as (1) 0–2 cups per day, (2) 3–4 cups per day, (3) 5–6 cups per day, (4) 7–9 cups per day and (5) ≥10 cups per day.⁶³ For ERF, which is a family-based cohort, the analysis was performed using a mixed model^{66,67} incorporating a relationship matrix estimated from the genotyped data.⁶⁸ A fixed-effects meta-analysis was conducted in METAL using the inverse variance-weighted method. All SNPs that had a low minor allele frequency (<0.01) and low imputation quality (Rsq/proper_info <0.3) were dropped from the meta-analysis. Genomic control correction was also applied to all cohorts before meta-analysis. Heterogeneity between studies was assessed with Cochran's Q-test in METAL.

Replication analysis

We sought *in silico* replication of the SNPs that showed a *P*-value <1 × 10⁻⁰⁶ (Table 1) in an independent

Table 1 Top hits of the meta-analysis

Marker name	Allele 1	Allele 2	β ^a	σ _β ^b	P-value	Direction ^c	Chr.	Position	MAF	RSQ	Gene	Feature
rs2470893	t	c	0.0614	0.011	2.39 × 10 ⁻⁰⁸	+	15	72806502	0.26–0.35	0.79–1	CYP1A1/ CYP1A2	Intergenic
rs2472297	t	c	0.0671	0.012	4.18 × 10 ⁻⁰⁸	+	15	72814933	0.17–0.31	0.56–1	CYP1A1/ CYP1A2	Intergenic
rs6495122	a	c	-0.0551	0.010	8.22 × 10 ⁻⁰⁸	---	15	72912698	0.36–0.46	0.93–1	CPLX3/ ULK3	Intergenic
rs16868941	a	g	0.0648	0.0124	1.55 × 10 ⁻⁰⁷	+	8	103121553	0.17–2.3	0.93–1	NCALD	Intron[NM_001040630.1]
rs382140	a	g	0.0729	0.0143	3.34 × 10 ⁻⁰⁷	+	7	107569436	0.11–0.21	0.61–1	LAMB4/ NRCAM	Intergenic
rs9526558	a	g	0.0595	0.012	6.79 × 10 ⁻⁰⁷	+	13	48880513	0.21–0.26	0.97–1	CAB39L	Intron[NM_001079670.1]

Abbreviations: Chr., chromosome; MAF, minor allele frequency.

^aThe effect estimate of the single-nucleotide polymorphism is shown.

^bs.e. of the effect estimate is shown.

^cDirection of the effect of allele 1 is shown. See Figure 2 for details.

Dutch cohort (LifeLines, $N=7929$, % of women = 57). The LifeLines Cohort Study was a multi-disciplinary prospective population-based cohort study examining in a unique three-generation design the health and health-related behaviors of 165 000 individuals living in the North East region of The Netherlands. All survey participants were between 18 and 90 years of age at the time of enrollment. Recruitment had been going on since the end of 2006, and until January 2011, >40 000 participants were included. Genotyping for LifeLines was performed on Illumina CytoSNP12 v2. Genetic imputations were performed in BEAGLE (Brian L. Browning, Seattle, WA, USA) v3.1.0 using build 36 HapMap CEU v22 as the reference population. Statistical analysis was performed in PLINK using the same analysis model as in the discovery phase. Random-effects meta-analysis was performed for the top hits in the 'rmeta' library of the R package.

Gene expression analysis

We used public gene expression databases for lymphocytes and the brain to examine the expression of genes implicated by the GWAS^{69–71} (<http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/>). In addition, we performed caffeine treatment on three different cell types as part of a project to find inherited protein-truncating mutations by Gene Inhibition of Nonsense Mediated Decay:⁷² (1) lymphoblastoid cell lines (LCLs) established from the blood of 24 females from hereditary non-*BRCA1/2* breast cancer families recruited into the Kathleen Cuninghame Foundation for Research into Breast Cancer (kConFab), (2) a cell line newly established from a breast-to-bone metastasis and (3) the colon cancer cell line, HT29. Experiments were performed in triplicate for each sample. Optimization experiments determined cell number and caffeine concentration for each cell type. For LCLs, 3.5×10^6 cells were plated in 10 ml of tissue culture medium (RPMI-1640 supplemented with 10% fetal calf serum) containing 7.5 mM caffeine (Sigma, St Louis, MO, USA). For both the HT29 and the bone metastasis cell line, 1×10^6 cells were plated in 10 ml of tissue culture medium (RPMI-1640 supplemented with 10% fetal calf serum) 24 hours before treating cells with media containing 10 mM caffeine (Sigma). All cell lines were incubated with caffeine for a total of 8 h at 37 °C in a humidified atmosphere containing 5% carbon dioxide. Matching untreated cells were used as controls.

Total RNA was extracted from both caffeine-treated and untreated control cells using the RNeasy RNA extraction kit (Qiagen, Hilden, Germany) as per the manufacturer's instructions. Biotinylated cRNA was prepared from 450 ng total RNA using the Illumina TotalPrep RNA amplification kit (Ambion, Foster City, CA, USA). After quantification using a ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, Delaware, USA), a total of 156 samples (750 ng cRNA per sample) were hybridized to Human HT-12 Expression BeadChips (Illumina) using all manufacturer's reagents for washing, detecting and

scanning as per the whole-genome gene expression direct hybridization assay protocol (Illumina).

The HumanHT-12 Expression BeadChips (Illumina) contains 48 803 probes that cover >25 000 annotated genes. Expression data were collated and quality checked in the Illumina BeadStudio and then imported into GeneSpring V10.0 (Agilent Technologies, Santa Clara, CA, USA). Data were quantile normalized to the baseline of the median of all samples and then filtered using an Illumina detection score of >0.95 in at least one sample. To identify differentially expressed genes between each caffeine-treated sample to its own untreated control, a linear model was implemented using R and the LIMMA (Smyth, 2004)⁷³ package. Data were adjusted for multiple testing with a false discovery rate of 5%. Two criteria were used to select the set of relevant genes. First, a set of genes with a log odds > 0 (B-statistic) was selected, followed by a hierarchical search of genes based on log-fold changes. Gene lists were then imported into GeneSpring V10.0 (Agilent Technologies) for data visualization and to determine overlapping genes between different cell types. To see which biological pathways are significantly enriched, a pathway analysis was performed using PANTHER⁷⁴ of all genes that showed significant differential expression after correction for multiple testing (P -value < Bonferroni's threshold = $0.05/48\,803 = 1 \times 10^{-06}$) in at least one cell type. An additional test was performed to test for enrichment of specific biological pathways or canonical functions in the Ingenuity Pathway Analysis program (Ingenuity Systems, Inc., Redwood City, CA, USA).

Results

Genome-wide association analysis

The descriptive statistics of each cohort are provided in Supplementary Table 2. Genome-wide association results are provided in Supplementary Figure 1, and the quantile–quantile plot is provided in the Supplementary Figure 2. In the classical GWAS, two SNPs rs2470893 and rs2472297 exceeded the genome-wide significance threshold of 5×10^{-08} , with the best hit rs2470893 ($P = 2.3 \times 10^{-08}$) (Table 1). The two SNPs are in strong linkage disequilibrium (LD) ($r^2 = 0.70$) and are located at 15q24, between the genes *CYP1A1* and *CYP1A2* (Figure 1). Significant heterogeneity between studies was detected at both loci (rs2470893, $P_{\text{HET}} = 0.01$ and rs2472297, $P_{\text{HET}} = 0.0008$) (Supplementary Table 3), although the direction of the effect for the top hit was consistent across all populations (Figure 2). Neither of the SNPs are in strong LD with rs762551, the SNP previously identified as increasing *CYP1A2* activity in caffeine-administered smokers ($r^2 = 0.12$ and $r^2 = 0.06$, respectively), and rs762551 showed only nominal significance in the analysis ($P = 0.003$). The minor 'T' allele of the most significant SNP rs2470893 was consistently positively associated with coffee consumption across all the cohorts with effect estimates ranging from 0.013 to 0.169 (Figure 2). rs2470893 was

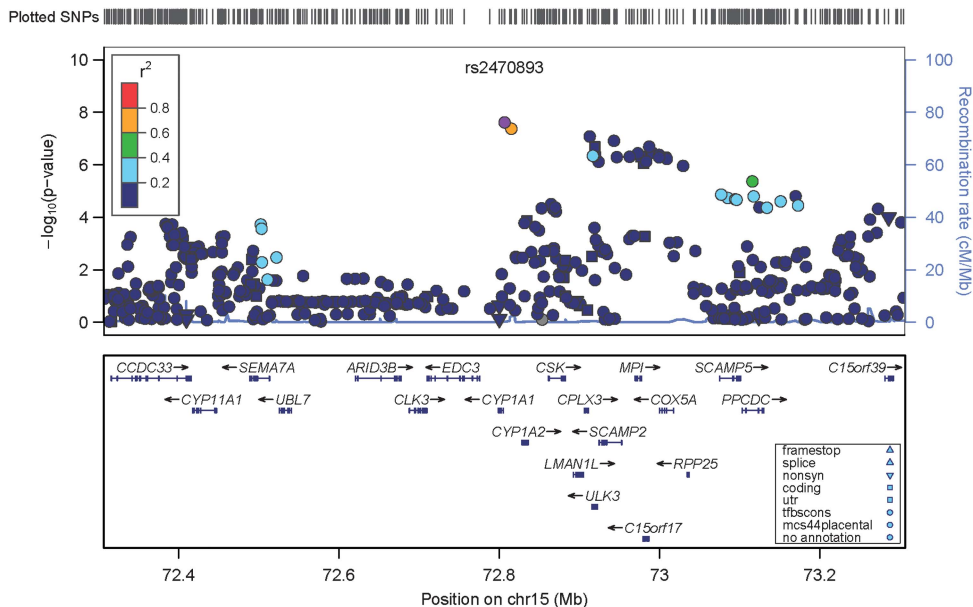


Figure 1 Regional association plot for 15q24. The vertical axis shows the negative logarithm of the association P -values, and the horizontal axis shows the position in mega bases. Each dot represents an SNP and the colors of the dots represent the extent of linkage disequilibrium with the top SNP, which is colored in purple. Genes in the region are shown below the horizontal axis.

genotyped in four cohorts, and in these, the minor allele frequency ranged from 0.26 to 0.35. Imputation quality was high in the other cohorts ranging from 0.79 to 1. The association signals at the *CYP1A1/CYP1A2* locus remain unchanged when a separate analysis adjusted for smoking status was performed in the two cohorts: ERF (rs2470893, P -value = 8.5×10^{-04} and rs2472297, P -value = 8.8×10^{-04}) and RS-II (rs2470893, P -value = 5.3×10^{-07} and rs2472297, P -value = 6.3×10^{-08}), which were the primary cohorts driving the association signal at 15q24.

Figure 1 shows that there are a number of signals in other genes in the chromosome 15 region. A strong association was observed for rs6495122 (P -value = 8.22×10^{-08}) (Table 1). This SNP is located at chromosome 15q24 in the intergenic region between the genes *ULK3* and *CPLX3* and does not appear to be in strong LD with the two best hits in this region (rs2470893, r^2 = 0.175 and rs2472297, r^2 = 0.086) (Figure 1). A secondary conditional association analysis of the 15q24 region in RS-II (which was the largest contributor to the association signal at 15q24, P -value = 8.4×10^{-06}) adjusting for the most significant finding (rs2470893) revealed nominal association of rs2472297 (P -value = 4.4×10^{-02}) and moderate association of rs6495122 (P -value = 1.1×10^{-03}) and rs12487 (P -value = 7.7×10^{-04}) with coffee drinking (Supplementary Figure 3).

Other SNPs on different chromosomes that showed strong evidence of association (P -values below 10^{-06}) are listed in Table 1. These include rs16868941, which is an intronic SNP in the *NCALD* gene (P -value = 1.5×10^{-07} ; Supplementary Figure 4); rs382140, an intergenic SNP between the genes *LAMB4*

and *NRCAM* (P -value = 3.3×10^{-07} ; Supplementary Figure 5) and rs9526558, an intronic variant within *CAB39L* (P -value = 6.79×10^{-07} ; Supplementary Figure 6). No heterogeneity was detected at any of these loci (Table 2).

Replication analysis

Table 2 provides results from replication analysis and the meta-analysis of discovery and replication cohorts. Replication was sought for six SNPs presented in Table 1, which necessitated a Bonferroni-corrected significance threshold of 8.3×10^{-03} . Owing to bad imputation quality in the replication cohort, no replication was performed for rs9526558. A significant association was observed in the replication analysis for the two genome-wide significant hits; rs2470893 (P -value = 7.2×10^{-05}) and rs2472297 (P -value = 1.9×10^{-05}) in the *CYP1A1/CYP1A2* region (Table 2, Figure 2). The meta-analysis of the discovery and replication cohorts showed a strongly significant association of rs2470893 (P -value = 1.6×10^{-11}) and rs2472297 (P -value = 2.7×10^{-11}). Among marginally significant SNPs in the genome-wide analysis, the SNP rs382140 near the *NRCAM* gene also showed significant association in the replication analysis (P -value = 1.3×10^{-03}). The meta-analysis of the discovery and replication cohorts yielded a genome-wide significant association of rs382140 with coffee drinking (P -value = 3.9×10^{-09}) (Table 2, Figure 2). rs6495122 also showed nominal significance in the replication analysis (P -value = 0.02). Although this P -value did not pass the Bonferroni threshold, the meta-analysis of discovery and replication cohorts showed genome-wide significant association of

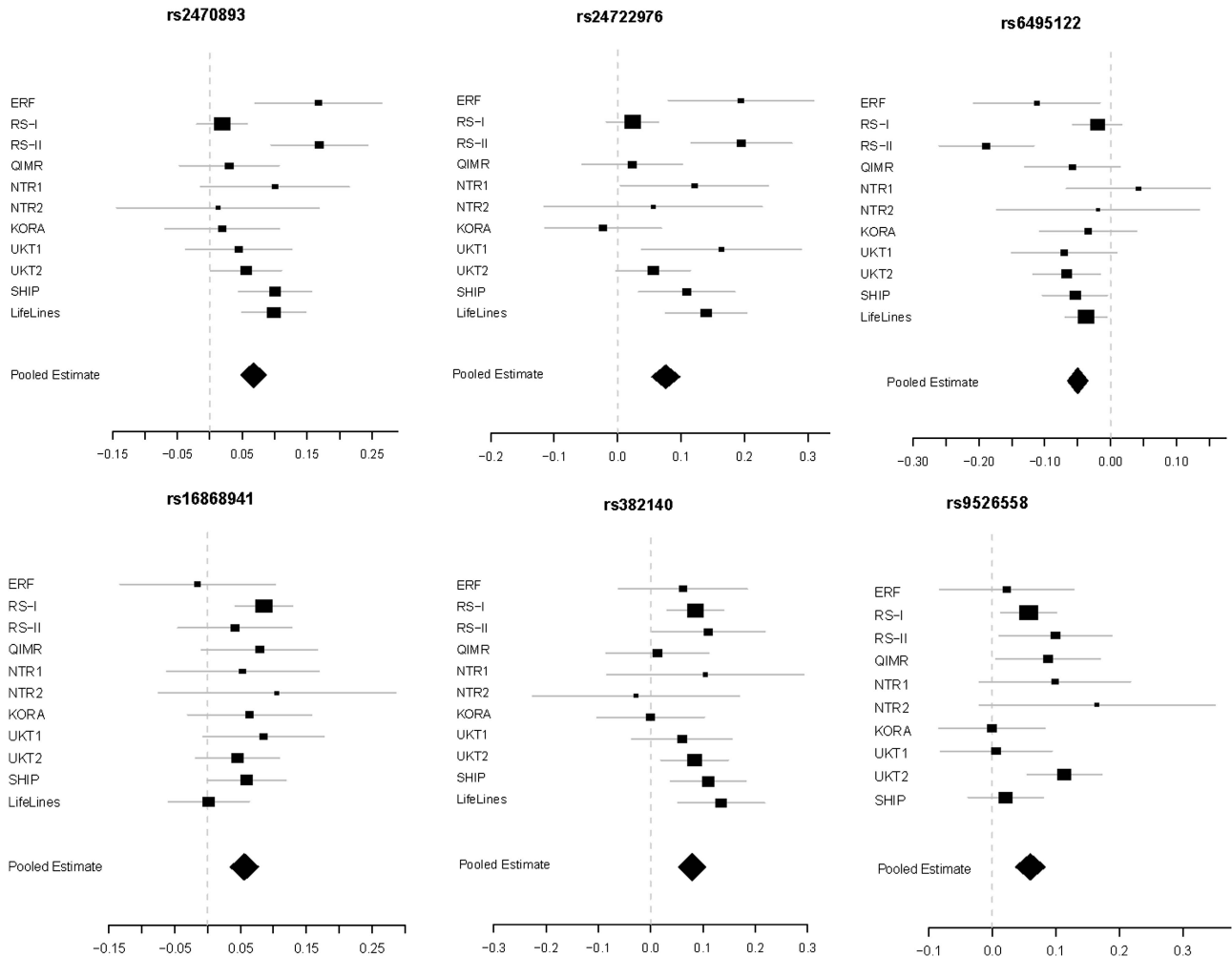


Figure 2 Forest plot for the top hits from the GWAS. Vertically left, the populations included in the GWAS and the replication phase. The boxes represent precision and horizontal lines represent the confidence intervals. The diamond represents the pooled effect estimate from the meta-analysis of all cohorts. The horizontal axis shows the scale of the effects.

rs6495122 with coffee drinking (P -value = 7.1×10^{-09}) (Table 2, Figure 2).

Additional evidence of association was sought from independent GWAS performed by the deCODE, including samples from Iceland ($n=3027$) and The Netherlands (Nijmegen) ($n=2812$). These studies performed GWAS on coffee consumption in cups per day only among coffee drinkers. Among the marginally significant findings in the GWAS, only the SNP rs9526558 in the *CAB39L* gene showed consistent direction and size of the effect estimate but an insignificant association in the deCODE samples (Supplementary Table 4). In contrast, the *CYP1A1/CYP1A2* region is significantly associated with coffee consumption in this data set (rs2472297, P -value = 5.4×10^{-14}),⁷⁵ underscoring a robust association at the *CYP1A1/CYP1A2* loci.

Expression analyses

First, we evaluated to what extent the SNPs above are associated with a differential expression in public

expression databases. For none of the SNPs discussed above, there was evidence for altered expression in the brain.⁷¹ Supplementary Table 5 shows that rs2470893 is involved in the expression of the *COX5A* gene (P -value = 1.2×10^{-04}). Moreover, the second locus in the region tagged by rs2470893 is associated with *COX5A* expression. Of the four regions that did not reach genome-wide significance, only the chromosome 7 region shows evidence of differential expression in lymphocytes. rs382140 is associated with the expression of the *SEMA3D* gene (P -value = 2.0×10^{-04}), a gene most likely involved in axonal guidance. However, this SNP is also associated with the expression of several other genes in *trans* including *SUCLG2*, *TOLLIP*, *NRG2* and *HFE*. Of these genes, *HFE* is most well known in coffee research, being one of the major genes involved in iron metabolism. The protective effect of coffee for risk of type II diabetes mellitus is suggested to be at least partially explained by the iron absorption inhibitory effect of coffee.^{76,77}

Table 2 Results of replication analysis

Marker name	Allele		Chr.	Position	Gene	Discovery (N = 18 176)		Replication (N = 7929)		Meta-analysis (discovery + replication) fixed effect				
	1	2				β	σ_β	P-value	β	σ_β	P-value	β	σ_β	P-value
rs2470893	t	c	15	72806502	CYP1A1/ CYP1A2	0.0614	0.011	2.39 × 10 ⁻⁰⁸	0.0986	0.024	7.2 × 10 ⁻⁰⁵	0.0675	0.010	1.6 × 10 ⁻¹¹
rs2472297	t	c	15	72814933	CYP1A1/ CYP1A2	0.0671	0.012	4.18 × 10 ⁻⁰⁸	0.1396	0.032	1.9 × 10 ⁻⁰⁵	0.076	0.011	2.7 × 10 ⁻¹¹
rs6495122	a	c	15	72912698	CPLX3/ ULK3	-0.0551	0.010	8.22 × 10 ⁻⁰⁸	-0.0375	0.016	0.020	-0.05	0.008	7.1 × 10 ⁻⁰⁹
rs16868941	a	g	8	103121553	NCALD	0.0648	0.0124	1.55 × 10 ⁻⁰⁷	0.0018	0.031	0.954	0.056	0.011	8.2 × 10 ⁻⁰⁷
rs382140	a	g	7	107569436	LAMB4/ NRCAM	0.0729	0.0143	3.34 × 10 ⁻⁰⁷	0.1351	0.042	1.4 × 10 ⁻⁰³	0.079	0.014	3.9 × 10 ⁻⁰⁹
rs9526558	a	g	13	48880513	CAB39L	0.0595	0.012	6.79 × 10 ⁻⁰⁷	NA	NA	NA	NA	NA	NA

Abbreviation: Chr., chromosome. Significant P-values are represented in bold.

Next, we examined to what extent caffeine alters expression of the genes, which are tagged by the associated SNPs reported above. A total of 647 autosomal genes were found to be differentially expressed in all three types of cell lines (data not shown). *CYP1A1* encoding the protein that metabolizes polycyclic aromatic hydrocarbons was found to be downregulated by caffeine in LCLs (fold-change = 1.29, $P\text{-value}_{\text{adj}} = 5.38 \times 10^{-08}$, $\beta = 9.54$), but no evidence of differential expression after caffeine treatment was observed for *CYP1A2* or the *NRCAM* genes in any of the three cell types. However, the enzyme encoded by the *CYP1A2* gene acts primarily in the liver as part of the cytochrome P450 system, and we did not treat any liver-derived cell lines. *CAB39L* was differentially expressed in all three cell types. The change in expression after caffeine treatment was most pronounced in LCLs (fold-change = 0.23, $P\text{-value}_{\text{adj}} = 3.8 \times 10^{-27}$, $\beta = 55.03$). In combination with GWAS results, with finding, this finding suggests that *CAB39L* is involved in caffeine metabolism and intake. There are two probes on the Illumina array that represent the *NRCAM* gene. However, neither of these probes was in the differentially expressed gene lists for all three cell line types. That is, gene expression of *NRCAM* was not found to be differentially expressed after treatment with caffeine.

In the pathway analysis, we included 4886 genes that were differentially expressed in at least 1 cell type and in total 147 pathways were tested (Supplementary Table 6). The pathway analysis suggests significant overrepresentation of genes in the ubiquitin proteasome pathway ($P\text{-value} = 2.2 \times 10^{-05}$), p53 ($P\text{-value} = 3.5 \times 10^{-05}$), Parkinson's disease ($P\text{-value} = 3.6 \times 10^{-05}$), purine biosynthesis ($P\text{-value} = 9 \times 10^{-05}$) and cell-cycle ($P\text{-value} = 1.4 \times 10^{-04}$) pathways. Genes involved in various metabolic processes were significantly overrepresented ($P\text{-value} = 5.6 \times 10^{-56}$).

Coffee-related phenotypes

We tested for the association of the top hits from the GWAS with two established coffee-related phenotypes: blood pressure and AD. For blood pressure and hypertension, we used the data from CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology). A significant association was found for the top hit rs2470893 with systolic ($P\text{-value} = 5 \times 10^{-03}$) and diastolic blood pressure ($P\text{-value} = 1.6 \times 10^{-04}$), and a borderline association was observed with hypertension ($P\text{-value} = 0.02$). The 'T' allele which was associated with increased coffee drinking was also associated with increased blood pressure. This association appears to be independent of rs6495122 (Figure 1, $r^2 = 0.175$), which was found to be associated with blood pressure in a previous study.⁷⁸ For AD, we used the data from RS and two previously published GWASs on AD.^{79,80} No association of any of the SNPs tested was detected with AD.

Discussion

We conducted a meta-analysis of GWAS on coffee consumption from 8 cohorts comprising >18 000 individuals of Northern European ancestry and attempted replication of our top findings from the GWAS in another ~8000 individuals from an independent cohort. Successful replication of the two genome-wide significant hits rs2470893 and rs2472297, which are also in strong LD and located between the *CYP1A1* and *CYP1A2* genes, and differential expression of the *CYP1A1* gene after caffeine treatment in LCLs strongly implicates the *CYP1A1/CYP1A2* locus in coffee drinking. Our study also suggests significant association of an SNP rs382140 in the promoter region of the *NRCAM* gene with coffee drinking, and the combined results from association (rs9526558) and expression analyses implicate *CAB39L* in coffee drinking. Although significant heterogeneity was detected at the *CYP1A1/CYP1A2* locus ($P_{\text{HET}}=0.01$) and the random effects model did not show genome-wide significance. Figure 2 highlights that the effect estimates were largely consistent across all populations. The *CYP1A1/CYP1A2* locus is also significantly implicated recently in a separate meta-analysis of GWAS of coffee drinking conducted by deCODE in ~6000 coffee drinkers⁷⁵ and a GWAS on caffeine intake⁸¹ further strengthening our data.

Caffeine is known to be metabolized primarily by *CYP1A2* in the liver, and hence *CYP1A2* has long been a candidate gene for coffee consumption. A previous study showed that C->A polymorphism in the intron I of *CYP1A2* was associated with increased caffeine metabolism.⁸² The *CYP1A2* gene encodes a P450 enzyme involved in O-de-ethylation of phenacetin.⁸³ The human hepatic microsomal caffeine 3-demethylation, which is the initial major step in caffeine biotransformation in humans, is selectively catalyzed by *CYP1A2*.⁸³ *CYP1A1* encodes the member P1-450 of the cytochrome P450 superfamily of enzymes. P1-450 is most closely associated with polycyclic-hydrocarbon-induced aryl hydrocarbon hydroxylase (AHH) activity and is known to metabolize polycyclic aromatic hydrocarbons such as benzo(a)pyrene, which is a constituent of coffee known to be involved in carcinogenesis.¹³ The *CYP1A1* and *CYP1A2* genes are separated by a 23-kb segment that contains no other open-reading frames. They are in opposite orientation, revealing that they share a common 5' flanking region.⁸⁴ Cytochrome P450 genes are a superfamily of heme-containing mono-oxygenases that metabolize many xenobiotics, including drugs, carcinogens and toxicants, as well as endogenous compounds such as fatty acids and neurotransmitters. *CYP1A1* has been identified in the human brain⁸⁵ and localized to the cortical regions, midbrain, basal ganglia and cerebellum,⁸⁶ whereas *CYP1A2* has been found in most brain regions examined.^{87,88}

A third SNP—rs6495122—approached genome-wide significance ($P=8.22 \times 10^{-06}$). This SNP is also

located on chromosome 15, but is not in strong LD with the genome-wide significant SNPs ($r^2=0.175$ and $r^2=0.086$, respectively). The secondary conditional analysis in RS-II suggests that an association signal at rs6495122 is partially independent of the top hit in this region (rs2470893). A previous meta-analysis found this SNP to be significantly associated with diastolic blood pressure, and nominally associated with systolic blood pressure and hypertension.⁷⁸ There is evidence that suggest that caffeine consumption is associated with an increase in blood pressure. Consuming a dose of caffeine equivalent to 2–3 cups of coffee (200–250 mg) has been found to increase both systolic and diastolic blood pressure.¹⁰ Results from studies examining the long-term effects of caffeine on blood pressure and risk to hypertension have been mixed.²¹ Our studies underscore a joint genetic background of coffee consumption and blood pressure.

Of the non-15q24 loci, rs382140 near the *NRCAM* is an interesting locus. This SNP showed a P -value of 3.9×10^{-09} in the association analysis. The SNP rs382140 did not show evidence of association in the deCODE sample. The reason may be the exclusion of non-drinkers in the deCODE sample as deCODE restricted the analysis to those who drink coffee. *NRCAM* encodes neuronal cell adhesion molecule, is expressed in the brain and is involved in several aspects of nervous system development. Allelic variants of this gene have been associated with autism⁸⁹ and addiction vulnerability.^{90,91} Genetic variants that increase the risk to addiction may also influence coffee consumption, as twin studies have shown that there is some overlap in the heritability of coffee consumption and use of nicotine and alcohol.^{92,93} Caffeine has been described as 'a model drug of abuse'^{26,94,95} and the finding of the *NRCAM* gene is in line with this hypothesis, although we did find *NRCAM* to be differentially expressed after caffeine treatment. The expression of *NRCAM* was found to be upregulated in papillary thyroid carcinomas.⁹⁶

CAB39L was found to be upregulated in all three cell types used in the expression analysis in this study in addition to strong association signals in the meta-analysis (rs9526558, P -value = 6.8×10^{-07}). The SNP rs9526558 was not available for replication because of bad imputation quality in the replication cohort; however, in deCODE, it showed consistency of the effect size and direction, although significance was not achieved. This may be attributed to the different study design and lower power of the deCODE sample. *CAB39L* encodes calcium-binding protein 39-like, a gene that is expressed in the brain. The function of this gene is not well characterized, but the encoded protein interacts with the serine/threonine kinase 11 (*STK11*) gene. This gene, which encodes a member of the serine/threonine kinase family, regulates cell polarity and functions as a tumor suppressor.

An intronic SNP in the gene *NCALD* also showed suggestive evidence of association with coffee

consumption (rs16868941 $P = 1.5 \times 10^{-07}$). Like *CAB39L*, *NCALD* is involved in calcium metabolism. *NCALD* encodes a member of the neuronal calcium sensor family of calcium-binding proteins. The protein is cytosolic at resting calcium levels but elevated intracellular calcium levels induce a conformational change that exposes the myristoyl group, resulting in protein association with membranes and partial co-localization with the perinuclear trans-golgi network. The protein is considered a regulator of G protein-coupled receptor signal transduction. Several alternatively spliced variants of this gene have been determined, all of which encode the same protein. The gene has shown to be associated with diabetic nephropathy.⁹⁷ Earlier we found evidence for association of the *NCALD* gene with sleep latency (rs17498920, P -value = 2×10^{-05}) (NA, unpublished data). Coffee intake is known to interfere with melatonin secretion,⁹⁸ thereby delaying the onset of sleep.

It is interesting to note that the pathway analysis implicates both ubiquitin proteasome and Parkinson's disease pathways as being significantly enriched with genes differentially expressed after caffeine treatment, given that coffee/caffeine is known to protect against Parkinson's disease.⁹⁹ Mutations in the *PARKIN* gene are the most common cause of hereditary Parkinsonism, and Parkin is an enzyme that ubiquitinates several candidate substrate proteins, thereby targeting them for proteasomal degradation.¹⁰⁰

In this study, we have combined gene expression data from three cell types treated with caffeine with GWAS data to study genes implicated in coffee-drinking habits in more detail. Studies that contributed to our GWAS came from different countries. Differences across studies and within a study, for instance cup size and caffeine content could have led to a loss of power in gene discovery. Significant heterogeneity between studies was observed at the *CYP1A1/CYP1A2* locus, although the effect estimates were consistent. To overcome the problem of heterogeneity, we also performed a random effects meta-analysis.

Our gene expression studies in lymphocytes strengthen the evidence of association for *CYP1A1* and *CAB39L*. Although in analyzing the effects of caffeine, it would be most pertinent to study changes in other cells and tissues including the liver and brain tissues, there is some evidence to suggest that gene expression in blood can be a good marker of gene expression in the central nervous system, and so LCLs may provide a good alternative.¹⁰¹ It should be noted that the concentration of caffeine we used is in the lower limit of that found in the bloodstream after drinking a cup of coffee. A recent study suggests that there are different patterns of gene expression at various concentrations of caffeine,²⁵ and further studies of gene expression should use different concentrations of caffeine within the normal range found after caffeine consumption in humans.

The fact that an individual would choose to abstain from coffee or drink only very small quantities may relate to insomnia, anxiety, trembling or other side effects of caffeine. It will likely be informative to test whether carriers of coffee consumption alleles experience the same or different side effects from caffeine. Analysis of individuals according to caffeine symptoms and compared with controls may provide more power for detecting caffeine-sensitive alleles. Our study is an important step in understanding the genetics of coffee/caffeine consumption. Caffeine is the most used psychoactive drug in the world, and coffee is the most common form of caffeine consumption among adults. Therefore, our results have implications not only for understanding individual differences in caffeine consumption but also for many other human traits and diseases such as blood pressure. Previous overlap in SNPs identified for caffeine-induced anxiety and panic disorder indicate that an understanding of how caffeine mediates its effects will help decipher the genetics of anxiety and anxiety disorders. Similarly, other studies have identified interactions between long-term caffeine use, common variants and risk of disease and our results can inform future studies in these areas.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

ERF: The genotyping for the ERF study was supported by EUROSPAN (European Special Populations Research Network) and the European Commission FP6 STRP grant (018947; LSHG-CT-2006-01947). The ERF study was further supported by grants from the Netherlands Organization for Scientific Research, Erasmus MC, the Centre for Medical Systems Biology (CMSB) and the Netherlands Brain Foundation (HersenStichting Nederland). We are grateful to all patients and their relatives, general practitioners and neurologists for their contributions and to P Veraart for her help in genealogy, Jeannette Vergeer for the supervision of the laboratory work and P Snijders for his help in data collection. We acknowledge Internationale Stichting Alzheimer Onderzoek (ISAO) and Hersenstichting Netherlands. *KORA*: The KORA Augsburg studies were financed by the Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, and supported by grants from the German Federal Ministry of Education and Research (BMBF). Part of this work was financed by the German National Genome Research Network (NGFN). Our research was supported within the Munich Center of Health Sciences (MC Health) as part of LMUinnovativ. *Netherlands Twin Register (NTR)*: We acknowledge support from: Genetic basis of anxiety and depression (904-61-090); Twin-family database for behavior

genomics studies (480-04-004); Genetic determinants of risk behavior in relation to alcohol use and alcohol use disorder: a developmental perspective (ZonMW Addiction 31160008), Center for Medical Systems Biology (NWO Genomics); Spinozapremie (SPI 56-464-14192); Neuroscience campus Amsterdam (NCA-VU); and Genetics of Mental Illness (ERC 230374). Genotyping was funded in part by the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health. JMV is financially supported by NWO (VENI 451-06-004). *QIMR*: We thank Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, as well as the Clinical Follow-Up Study (funded 2001–2009 by NHMRC and currently by the National Breast Cancer Foundation and Cancer Australia no. 628333) for their contributions to this resource, and the many families who contribute to kConFab. kConFab is supported by grants from the National Breast Cancer Foundation, the National Health and Medical Research Council (NHMRC) and by the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia. We thank the twins and their families for their participation. We also thank Ann Eldridge, Marlene Grace, Kerrie McAloney (sample collection); Lisa Bowdler, Steven Crooks (DNA processing); Dale Nyholt, Sarah Medland and Scott Gordon (imputation and genotyping QC); Allan McRae and Peter Visscher for helpful advice and input. Funding was provided by the Australian National Health and Medical Research Council (241944, 339462, 389927, 389875, 389891, 389892, 389938, 442915, 442981, 496739, 552485, 552498), the Australian Research Council (A7960034, A79906588, A79801419, DP0770096, DP0212016, DP0343921), the FP-5 GenomeUtwinn Project (QLG2-CT-2002-01254) and the US National Institutes of Health (NIH grants AA07535, AA10248, AA13320, AA13321, AA13326, AA14041, MH66206). A portion of the genotyping on which this study was based (Illumina 370K scans on 4300 individuals) was carried out at the Center for Inherited Disease Research, Baltimore (CIDR), through an access award to our late colleague Dr Richard Todd (Psychiatry, Washington University School of Medicine, St Louis). Statistical analyses were carried out on the Genetic Cluster Computer, which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003). GWM and GCT are supported by the National Health and Medical Research Council (NHMRC) Fellowship Scheme. JJ was financially supported by the Cancer Council Queensland. *Rotterdam Study (RS-I and RS-II)*: The Rotterdam Study is supported by the Erasmus Medical Center and Erasmus University Rotterdam, The Netherlands Organization for Scientific Research (NWO), The Netherlands Organization for Health Research and Development (ZonMw), The Research Institute for Diseases in the Elderly (RIDE), The Netherlands Genomics Initiative,

the Ministry of Education, Culture and Science, the Ministry of Health, Welfare and Sports, the European Commission (DG XII) and the Municipality of Rotterdam. The contribution of inhabitants, general practitioners and pharmacists of the Ommoord district to the Rotterdam Study is gratefully acknowledged. AI was financially supported by the Internationaal Parkinson Fonds. *The Study of Health in Pomerania (SHIP)*: SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants nos 01ZZ9603, 01ZZ0103 and 01ZZ0403), the Ministry of Cultural Affairs, as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania. Genome-wide data have been supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany, and the Federal State of Mecklenburg, West Pomerania. The University of Greifswald is a member of the 'Center of Knowledge Interchange' program of the Siemens AG. HJG was supported by the German Research Foundation; the Federal Ministry of Education and Research Germany; speakers honoraria from Bristol-Myers Squibb, Eli Lilly, Novartis, Eisai, Wyeth, Pfizer, Boehringer Ingelheim, Servier and travel funds from Janssen-Cilag, Lundbeck, Eli Lilly, Novartis, AstraZeneca and SALUS-Institute for Trend-Research and Therapy Evaluation in Mental Health. *Twins UK*: The study was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007–2013)/grant agreement HEALTH-F2-2008-201865-GEFOS and (FP7/2007–2013), ENGAGE project grant agreement HEALTH-F4-2007-201413 and the FP-5 GenomeUtwinn Project (QLG2-CT-2002-01254). The study also receives support from the Department of Health through the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London. TDS is an NIHR senior Investigator. The project also received support from a *Biotechnology and Biological Sciences Research Council* (BBSRC) project grant. (G20234). We acknowledge the funding and support of the National Eye Institute through an NIH/CIDR genotyping project (PI: Terri Young). We thank the staff from the Genotyping Facilities at the Wellcome Trust Sanger Institute for sample preparation, Quality Control and Genotyping led by Leena Peltonen and Panos Deloukas; Le Centre National de Génotypage, France, led by Mark Lathrop, for genotyping; Duke University, NC, USA, led by David Goldstein, for genotyping; and the Finnish Institute of Molecular Medicine, Finnish Genome Center, University of Helsinki, led by Aarno Palotie. *LifeLines*: The LifeLines Cohort Study, and generation and management of GWAS genotype data for the LifeLines Cohort Study is supported by the Netherlands Organization of Scientific Research NWO (grant 175.010.2007.006), the Economic Structure Enhancing Fund (FES) of the

Dutch government, the Ministry of Economic Affairs, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the Northern Netherlands Collaboration of Provinces (SNN), the Province of Groningen, University Medical Center Groningen, the University of Groningen, Dutch Kidney Foundation and Dutch Diabetes Research Foundation. We thank Behrooz Alizadeh, Annemieke Boesjes, Marcel Bruinenberg, Noortje Festen, Ilja Nolte, Lude Franke and Mitra Valimohammadi for their help in creating the GWAS database, and Rob Bieringa, Joost Keers, René Oostergo, Rosalie Visser and Judith Vonk for their work related to data-collection and validation. We are grateful to the study participants, the staff from the LifeLines Cohort Study and Medical Biobank Northern Netherlands, and the participating general practitioners and pharmacists.

References

- Ruusunen A, Lehto SM, Tolmunen T, Mursu J, Kaplan GA, Voutilainen S. Coffee, tea and caffeine intake and the risk of severe depression in middle-aged Finnish men: the Kuopio Ischaemic Heart Disease Risk Factor Study. *Public Health Nutr* 2010; **13**: 1215–1220.
- Sugiyama K, Kuriyama S, Akhter M, Kakizaki M, Nakaya N, Ohmori-Matsuda K et al. Coffee consumption and mortality due to all causes, cardiovascular disease, and cancer in Japanese women. *J Nutr* 2010; **140**: 1007–1013.
- Choi HK, Curhan G. Coffee consumption and risk of incident gout in women: the Nurses' Health Study. *Am J Clin Nutr* 2010; **92**: 922–927.
- Gongora-Alfaro JL. [Caffeine as a preventive drug for Parkinson's disease: epidemiologic evidence and experimental support] La cafeina como un farmaco preventivo de la enfermedad de Parkinson: evidencias epidemiológicas y sustrato experimental. *Rev Neurol* 2010; **50**: 221–229.
- van Dieren S, Uiterwaal CS, van der Schouw YT, van der AD, Boer JM, Spijkerman A et al. Coffee and tea consumption and risk of type 2 diabetes. *Diabetologia* 2009; **52**: 2561–2569.
- Trichopoulos D, Papapostolou M, Polychronopoulou A. Coffee and ovarian cancer. *Int J Cancer* 1981; **28**: 691–693.
- MacMahon B, Yen S, Trichopoulos D, Warren K, Nardi G. Coffee and cancer of the pancreas. *N Engl J Med* 1981; **304**: 630–633.
- Silvera SA, Jain M, Howe GR, Miller AB, Rohan TE. Intake of coffee and tea and risk of ovarian cancer: a prospective cohort study. *Nutr Cancer* 2007; **58**: 22–27.
- Jee SH, He J, Whelton PK, Suh I, Klag MJ. The effect of chronic coffee drinking on blood pressure: a meta-analysis of controlled clinical trials. *Hypertension* 1999; **33**: 647–652.
- Nurminen ML, Niittynen L, Korpela R, Vapaatalo H. Coffee, caffeine and blood pressure: a critical review. *Eur J Clin Nutr* 1999; **53**: 831–839.
- Cornelis MC, El-Sohehy A, Kabagambe EK, Campos H. Coffee, CYP1A2 genotype, and risk of myocardial infarction. *JAMA* 2006; **295**: 1135–1141.
- Kalhoff S, Ehmer U, Freiberg N, Manns MP, Strassburg CP. Coffee induces expression of glucuronosyltransferases by the aryl hydrocarbon receptor and Nrf2 in liver and stomach. *Gastroenterology* 2010; **139**: 1699–1710, 1710 e1–2.
- Nehlig A, Debry G. Potential genotoxic, mutagenic and antimutagenic effects of coffee: a review. *Mutat Res* 1994; **317**: 145–162.
- Kruijff N, Schouten A, Van der Stegen GHD. Occurrence of benzo[a]pyrene in roasted coffee, instant coffee and coffee brew. *The Cafe Cacao* 1987; **35**: 151–154.
- Nagao M, Takahashi Y, Yamanaka H, Sugimura T. Mutagens in coffee and tea. *Mutat Res* 1979; **68**: 101–106.
- Aeschbacher HU, Wolleb U, Loliger J, Spadone JC, Liardon R. Contribution of coffee aroma constituents to the mutagenicity of coffee. *Food Chem Toxicol* 1989; **27**: 227–232.
- Nukaya H, Iwami T, Ishida H, Tsuji K, Suwa Y, Wakabayashi K et al. N-2 acetylation of 2'-deoxyguanosine by coffee mutagens, methylglyoxal and hydrogen peroxide. *Mutat Res* 1990; **245**: 251–257.
- Krul C, Hageman G. Analysis of urinary caffeine metabolites to assess biotransformation enzyme activities by reversed-phase high-performance liquid chromatography. *J Chromatogr B Biomed Sci Appl* 1998; **709**: 27–34.
- Crews HM, Olivier L, Wilson LA. Urinary biomarkers for assessing dietary exposure to caffeine. *Food Addit Contam* 2001; **18**: 1075–1087.
- Selbach O, Haas HL. Hypocretins: the timing of sleep and waking. *Chronobiol Int* 2006; **23**: 63–70.
- Higdon JV, Frei B. Coffee and health: a review of recent human research. *Crit Rev Food Sci Nutr* 2006; **46**: 101–123.
- Fredholm BB, Battig K, Holmen J, Nehlig A, Zvartau EE. Actions of caffeine in the brain with special reference to factors that contribute to its widespread use. *Pharmacol Rev* 1999; **51**: 83–133.
- Cornelis MC, El-Sohehy A, Campos H. Genetic polymorphism of the adenosine A2A receptor is associated with habitual caffeine consumption. *Am J Clin Nutr* 2007; **86**: 240–244.
- Ferre S, von Euler G, Johansson B, Fredholm BB, Fuxe K. Stimulation of high-affinity adenosine A2 receptors decreases the affinity of dopamine D2 receptors in rat striatal membranes. *Proc Natl Acad Sci USA* 1991; **88**: 7238–7241.
- Yu L, Coelho JE, Zhang X, Fu Y, Tillman A, Karaoz U et al. Uncovering multiple molecular targets for caffeine using a drug target validation strategy combining A2A receptor knockout mice with microarray profiling. *Physiol Genomics* 2009; **37**: 199–210.
- Holtzman SG. Caffeine as a model drug of abuse. *Trends Pharmacol Sci* 1990; **11**: 355–356.
- Juliano LM, Griffiths RR. A critical review of caffeine withdrawal: empirical validation of symptoms and signs, incidence, severity, and associated features. *Psychopharmacology (Berl)* 2004; **176**: 1–29.
- Luciano M, Kirk KM, Heath AC, Martin NG. The genetics of tea and coffee drinking and preference for source of caffeine in a large community sample of Australian twins. *Addiction* 2005; **100**: 1510–1517.
- Mitchell PJ, Redman JR. Effects of caffeine, time of day and user history on study-related performance. *Psychopharmacology (Berl)* 1992; **109**: 121–126.
- Silverman K, Mumford GK, Griffiths RR. Enhancing caffeine reinforcement by behavioral requirements following drug ingestion. *Psychopharmacology (Berl)* 1994; **114**: 424–432.
- Horne JA, Reyner LA. Counteracting driver sleepiness: effects of napping, caffeine, and placebo. *Psychophysiology* 1996; **33**: 306–309.
- James JE. Does caffeine enhance or merely restore degraded psychomotor performance? *Neuropsychobiology* 1994; **30**: 124–125.
- James JE. Caffeine and psychomotor performance revisited. *Neuropsychobiology* 1995; **31**: 202–203.
- Rogers PJ, Richardson NJ, Derroncourt C. Caffeine use: is there a net benefit for mood and psychomotor performance? *Neuropsychobiology* 1995; **31**: 195–199.
- Christopher G, Sutherland D, Smith A. Effects of caffeine in non-withdrawn volunteers. *Hum Psychopharmacol* 2005; **20**: 47–53.
- Evans SM, Griffiths RR. Dose-related caffeine discrimination in normal volunteers: individual differences in subjective effects and self-reported cues. *Behav Pharmacol* 1991; **2**: 345–356.
- Childs E, Hohoff C, Deckert J, Xu K, Badner J, de Wit H. Association between ADORA2A and DRD2 polymorphisms and caffeine-induced anxiety. *Neuropsychopharmacology* 2008; **33**: 2791–2800.
- Klatsky AL, Armstrong MA, Friedman GD. Coffee, tea, and mortality. *Ann Epidemiol* 1993; **3**: 375–381.
- Vink JM, Staphorsius AS, Boomsma DI. A genetic analysis of coffee consumption in a sample of Dutch twins. *Twin Res Hum Genet* 2009; **12**: 127–131.

- 40 Laitala VS, Kaprio J, Silventoinen K. Genetics of coffee consumption and its stability. *Addiction* 2008; **103**: 2054–2061.
- 41 Gu L, Gonzalez FJ, Kalow W, Tang BK. Biotransformation of caffeine, paraxanthine, theobromine and theophylline by cDNA-expressed human CYP1A2 and CYP2E1. *Pharmacogenetics* 1992; **2**: 73–77.
- 42 Kalow W, Tang BK. The use of caffeine for enzyme assays: a critical appraisal. *Clin Pharmacol Ther* 1993; **53**: 503–514.
- 43 Rasmussen BB, Brix TH, Kyvik KO, Brosen K. The interindividual differences in the 3-demethylation of caffeine alias CYP1A2 is determined by both genetic and environmental factors. *Pharmacogenetics* 2002; **12**: 473–478.
- 44 Palatini P, Ceolotto G, Ragazzo F, Dorigatti F, Saladini F, Papparella I et al. CYP1A2 genotype modifies the association between coffee intake and the risk of hypertension. *J Hypertens* 2009; **27**: 1594–1601.
- 45 Kotsopoulos J, Ghadirian P, El-Sohemy A, Lynch HT, Snyder C, Daly M et al. The CYP1A2 genotype modifies the association between coffee consumption and breast cancer risk among BRCA1 mutation carriers. *Cancer Epidemiol Biomarkers Prev* 2007; **16**: 912–916.
- 46 Sachse C, Brockmoller J, Bauer S, Roots I. Functional significance of a C->A polymorphism in intron 1 of the cytochrome P450 CYP1A2 gene tested with caffeine. *Br J Clin Pharmacol* 1999; **47**: 445–449.
- 47 Deckert J, Nothen MM, Franke P, Delmo C, Fritze J, Knapp M et al. Systematic mutation screening and association study of the A1 and A2a adenosine receptor genes in panic disorder suggest a contribution of the A2a gene to the development of disease. *Mol Psychiatry* 1998; **3**: 81–85.
- 48 Hamilton SP, Slager SL, De Leon AB, Heiman GA, Klein DF, Hodge SE et al. Evidence for genetic linkage between a polymorphism in the adenosine 2A receptor and panic disorder. *Neuropsychopharmacology* 2004; **29**: 558–565.
- 49 Aulchenko YS, Heutink P, Mackay I, Bertoli-Avella AM, Pullen J, Vaessen N et al. Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet* 2004; **12**: 527–534.
- 50 Wichmann H, Gieger C, Illig T, Group M. KORA-gen—resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* 2005; **67**(Suppl 1): S26–S30.
- 51 Hofman A, Breteler MM, van Duijn CM, Janssen HL, Krestin GP, Kuipers EJ et al. The Rotterdam Study: 2010 objectives and design update. *Eur J Epidemiol* 2009; **24**: 553–572.
- 52 Boomsma DI, de Geus EJ, Vink JM, Stubbe JH, Distel MA, Hottenga JJ et al. Netherlands twin register: from twins to twin families. *Twin Res Hum Genet* 2006; **9**: 849–857.
- 53 Vink JM, Boomsma DI. A comparison of early and late respondents in a twin-family survey study. *Twin Res Hum Genet* 2008; **11**: 165–173.
- 54 John U, Greiner B, Hensel E, Ludemann J, Piek M, Sauer S et al. Study of Health In Pomerania (SHIP): a health examination survey in an east German region: objectives and design. *Soz Präventivmed* 2001; **46**: 186–194.
- 55 Volzke H, Alte D, Schmidt CO, Radke D, Lorbeer R, Friedrich N et al. Cohort profile: the study of health in pomerania. *Int J Epidemiol* 2010; **40**: 294–307.
- 56 Andrew T, Hart DJ, Snieder H, de Lange M, Spector TD, MacGregor AJ. Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Res* 2001; **4**: 464–477.
- 57 Spector TD, Williams FM. The UK adult twin registry (TwinsUK). *Twin Res Hum Genet* 2006; **9**: 899–906.
- 58 Medland SE, Nyholt DR, Painter JN, McEvoy BP, McRae AF, Zhu G et al. Common variants in the trichohyalin gene are associated with straight hair in Europeans. *Am J Hum Genet* 2009; **85**: 750–755.
- 59 Li Y, Abecasis GR. Rapid haplotype reconstruction and missing genotype inference. *Am J Hum Genet* 2006; **S79**: 2290.
- 60 Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annu Rev Genomics Hum Genet* 2009; **10**: 387–406.
- 61 Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009; **5**: e1000529.
- 62 Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007; **39**: 906–913.
- 63 Tuomilehto J, Hu G, Bidel S, Lindstrom J, Jousilahti P. Coffee consumption and risk of type 2 diabetes mellitus among middle-aged Finnish men and women. *JAMA* 2004; **291**: 1213–1219.
- 64 Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* 2010; **11**: 134.
- 65 Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; **447**: 661–678.
- 66 Chen WM, Abecasis GR. Family-based association tests for genome-wide association scans. *Am J Hum Genet* 2007; **81**: 913–926.
- 67 Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 2007; **23**: 1294–1296.
- 68 Amin N, van Duijn CM, Aulchenko YS. A genomic background based method for association analysis in related individuals. *PLoS One* 2007; **2**: e1274.
- 69 Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C et al. Population genomics of human gene expression. *Nat Genet* 2007; **39**: 1217–1224.
- 70 Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J et al. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* 2010; **464**: 773–777.
- 71 Myers AJ, Gibbs JR, Webster JA, Rohrer K, Zhao A, Marlowe L et al. A survey of genetic human cortical gene expression. *Nat Genet* 2007; **39**: 1494–1499.
- 72 Ivanov I, Lo KC, Hawthorn L, Cowell JK, Ionov Y. Identifying candidate colon cancer tumor suppressor genes using inhibition of nonsense-mediated mRNA decay in colon cancer cells. *Oncogene* 2007; **26**: 2873–2884.
- 73 Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* 2004; **3**.
- 74 Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, Thomas PD. PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res* 2010; **38**: D204–D210.
- 75 Sulem P, Gudbjartsson DF, Geller F, Prokopenko I, Feenstra B, Aben KK et al. Sequence variants at CYP1A1-CYP1A2 and AHR associate with coffee consumption. *Hum Mol Genet* 2011; **20**: 2071–2077.
- 76 Mascitelli L, Pezzetta F, Sullivan JL. Inhibition of iron absorption by coffee and the reduced risk of type 2 diabetes mellitus. *Arch Intern Med* 2007; **167**: 204–205; author reply 5.
- 77 Pereira MA, Parker ED, Folsom AR. Coffee consumption and risk of type 2 diabetes mellitus: an 11-year prospective study of 28 812 postmenopausal women. *Arch Intern Med* 2006; **166**: 1311–1316.
- 78 Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A et al. Genome-wide association study of blood pressure and hypertension. *Nat Genet* 2009; **41**: 677–687.
- 79 Lambert JC, Heath S, Even G, Campion D, Sleegers K, Hiltunen M et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* 2009; **41**: 1094–1099.
- 80 Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* 2009; **41**: 1088–1093.
- 81 Cornelis MC, Monda KL, Yu K, Paynter N, Azzato EM, Bennett SN et al. Genome-wide meta-analysis identifies regions on 7p21 (AHR) and 15q24 (CYP1A2) as determinants of habitual caffeine consumption. *PLoS Genet* 2011; **7**: e1002033.
- 82 Castorena-Torres F, Mendoza-Cantu A, de Leon MB, Cisneros B, Zapata-Perez O, Lopez-Carrillo L et al. CYP1A2 phenotype and genotype in a population from the Carboniferous Region of Coahuila, Mexico. *Toxicol Lett* 2005; **156**: 331–339.
- 83 Butler MA, Iwasaki M, Guengerich FP, Kadlubar FF. Human cytochrome P-450PA (P-450IA2), the phenacetin O-deethylase, is primarily responsible for the hepatic 3-demethylation of caffeine

- and N-oxidation of carcinogenic arylamines. *Proc Natl Acad Sci USA* 1989; **86**: 7696–7700.
- 84 Corchero J, Pimprale S, Kimura S, Gonzalez FJ. Organization of the CYP1A cluster on human chromosome 15: implications for gene regulation. *Pharmacogenetics* 2001; **11**: 1–6.
- 85 Yun CH, Park HJ, Kim SJ, Kim HK. Identification of cytochrome P450 1A1 in human brain. *Biochem Biophys Res Commun* 1998; **243**: 808–810.
- 86 McFayden MC, Melvin WT, Murray GI. Regional distribution of individual forms of cytochrome P450 mRNA in normal adult human brain. *Biochem Pharmacol* 1998; **55**: 825–830.
- 87 Morse DC, Stein AP, Thomas PE, Lowndes HE. Distribution and induction of cytochrome P450 1A1 and 1A2 in rat brain. *Toxicol Appl Pharmacol* 1998; **152**: 232–239.
- 88 Farin FM, Omiecinski CJ. Regiospecific expression of cytochrome P-450s and microsomal epoxide hydrolase in human brain tissue. *J Toxicol Environ Health* 1993; **40**: 317–335.
- 89 Sakurai T, Ramoz N, Reichert JG, Corwin TE, Kryzak L, Smith CJ et al. Association analysis of the NrCAM gene in autism and in subsets of families with severe obsessive-compulsive or self-stimulatory behaviors. *Psychiatr Genet* 2006; **16**: 251–257.
- 90 Ishiguro H, Liu QR, Gong JP, Hall FS, Ujike H, Morales M et al. NrCAM in addiction vulnerability: positional cloning, drug-regulation, haplotype-specific expression, and altered drug reward in knockout mice. *Neuropsychopharmacology* 2006; **31**: 572–584.
- 91 Matzel LD, Babiarz J, Townsend DA, Grossman HC, Grumet M. Neuronal cell adhesion molecule deletion induces a cognitive and behavioral phenotype reflective of impulsivity. *Genes Brain Behav* 2008; **7**: 470–480.
- 92 Swan GE, Carmelli D, Rosenman RH, Fabsitz RR, Christian JC. Smoking and alcohol consumption in adult male twins: genetic heritability and shared environmental influences. *J Subst Abuse* 1990; **2**: 39–50.
- 93 Swan GE, Carmelli D, Cardon LR. The consumption of tobacco, alcohol, and coffee in Caucasian male twins: a multivariate genetic analysis. *J Subst Abuse* 1996; **8**: 19–31.
- 94 Strain EC, Griffiths RR. Caffeine dependence: fact or fiction? *J R Soc Med* 1995; **88**: 437–440.
- 95 Hughes JR, Oliveto AH, Liguori A, Carpenter J, Howard T. Endorsement of DSM-IV dependence criteria among caffeine users. *Drug Alcohol Depend* 1998; **52**: 99–107.
- 96 Gorka B, Skubis-Zegadlo J, Mikula M, Bardadin K, Paliczka E, Czarnocka B. NrCAM, a neuronal system cell-adhesion molecule, is induced in papillary thyroid carcinomas. *Br J Cancer* 2007; **97**: 531–538.
- 97 Kamiyama M, Kobayashi M, Araki S, Iida A, Tsunoda T, Kawai K et al. Polymorphisms in the 3' UTR in the neurocalcin delta gene affect mRNA stability, and confer susceptibility to diabetic nephropathy. *Hum Genet* 2007; **122**: 397–407.
- 98 Shilo L, Sabbah H, Hadari R, Kovatz S, Weinberg U, Dolev S et al. The effects of coffee consumption on sleep and melatonin secretion. *Sleep Med* 2002; **3**: 271–273.
- 99 Ascherio A, Zhang SM, Hernan MA, Kawachi I, Colditz GA, Speizer FE et al. Prospective study of caffeine consumption and risk of Parkinson's disease in men and women. *Ann Neurol* 2001; **50**: 56–63.
- 100 Kahle PJ, Haass C. How does parkin ligate ubiquitin to Parkinson's disease? *EMBO Rep* 2004; **5**: 681–685.
- 101 Sullivan PF, Fan C, Perou CM. Evaluating the comparability of gene expression in blood and brain. *Am J Med Genet B Neuropsychiatr Genet* 2006; **141B**: 261–268.



This work is licensed under the Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>

Supplementary Information accompanies the paper on the Molecular Psychiatry website (<http://www.nature.com/mp>)