Sequence variants at CYP1A1–CYP1A2 and AHR associate with coffee consumption

Patrick Sulem[1],[*],[†], Daniel F. Gudbjartsson[1],[†], Frank Geller[2], Inga Prokopenko[3],[4],
Bjarke Feenstra[2], Katja K.H. Aben[5],[6], Barbara Franke[7],[8], Martin den Heijer[6],[9], Peter Kovacs[10],
Michael Stumvoll[11],[12], Reedik Mägi[3],[4], Lisa R. Yanek[13], Lewis C. Becker[13], Heather A. Boyd[2],
Simon N. Stacey[1], G. Bragi Walters[1], Adalbjorg Jonasdottir[1], Gudmar Thorleifsson[1],
Hilma Holm[1], Sigurjon A. Gudjonsson[1], Thorunn Rafnar[1], Gyda Björnsdottir[1],
Diane M. Becker[12], Mads Melbye[2], Augustine Kong[1], Anke Tönjes[11], Thorgeir Thorgeirsson[1],
Unnur Thorsteinsdottir[1],[14], Lambertus A. Kiemeney[5],[6],[15] and Kari Stefansson[1],[14]

[1]deCODE Genetics, Sturlugata 8, 101 Reykjavik, Iceland, [2]Department of Epidemiology Research, Statens Serum
Institut, 2300 Copenhagen S, Denmark, [3]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of
Oxford, Oxford OX3 7LJ, UK, [4]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN,
UK, [5]Department of Genetics, Comprehensive Cancer Center IKO, Nijmegen 6501 BG, The Netherlands,
[6]Department of Epidemiology, Biostatistics & HTA, [7]Department of Genetics, [8]Department of Psychiatry and
[9]Department of Endocrinology, Radboud University Nijmegen Medical Centre, Nijmegen 6500 HB, The Netherlands,
[10]Interdisciplinary Centre for Clinical Research, [11]Department of Medicine and [12]LIFE Study Center, University of
Leipzig, 04103 Leipzig, Germany, [13]The Johns Hopkins GeneSTAR Research Program, Division of General Internal
Medicine, The Johns Hopkins Medical Institutions, Baltimore, MD 21287, USA, [14]University of Iceland, Faculty of
Medicine, 101 Reykjavik, Iceland and [15]Department of Urology, Radboud University Nijmegen Medical Centre,
Nijmegen, 6500 HB, The Netherlands

Coffee is the most commonly used stimulant and caffeine is its main psychoactive ingredient. The heritability of coffee consumption has been estimated at around 50%. We performed a meta-analysis of four genome-wide association studies of coffee consumption among coffee drinkers from Iceland ($n = 2680$), the Netherlands ($n = 2791$), the Sorbs Slavonic population isolate in Germany ($n = 771$) and the USA ($n = 369$) using both directly genotyped and imputed single nucleotide polymorphisms (SNPs) (2.5 million SNPs). SNPs at the two most significant loci were also genotyped in a sample set from Iceland ($n = 2430$) and a Danish sample set consisting of pregnant women ($n = 1620$). Combining all data, two sequence variants significantly associated with increased coffee consumption: rs2472297-T located between *CYP1A1* and *CYP1A2* at 15q24 ($P = 5.4 \cdot 10^{-14}$) and rs6968865-T near aryl hydrocarbon receptor (*AHR*) at 7p21 ($P = 2.3 \cdot 10^{-11}$). An effect of ∼0.2 cups a day per allele was observed for both SNPs. CYP1A2 is the main caffeine metabolizing enzyme and is also involved in drug metabolism. *AHR* detects xenobiotics, such as polycyclic aryl hydrocarbons found in roasted coffee, and induces transcription of *CYP1A1* and *CYP1A2*. The association of these SNPs with coffee consumption was present in both smokers and non-smokers.

*To whom correspondence should be addressed. Tel: +354 5702371; Fax: +354 5701903; Email: patrick.sulem@decode.is
†These authors contributed equally to this work.

INTRODUCTION

Coffee is the most commonly used stimulant and caffeine is its main psychoactive ingredient (1). In many western countries, coffee is the main source of caffeine intake (1). Coffee tolerance and coffee withdrawal syndromes are common and linked to the antagonist effect of caffeine on the adenosine receptor (2). The effects of coffee consumption on health have not been firmly established. An increased risk of spontaneous abortion (3), cardiovascular disease (1), hypertension (4) and cancer (1) have been suggested, but also decreased risk of type 2 diabetes (1), Alzheimer's disease (5) and suicide (6). Twin studies have consistently estimated the heritability of coffee consumption to be ∼50% (7–9).

RESULTS

We performed a meta-analysis of genome-wide association studies of coffee consumption among coffee drinkers from four populations of European descent; from Iceland ($n = 2680$), the Netherlands ($n = 2791$), the Sorbs Slavonic population isolate in Germany ($n = 771$) and the USA ($n = 369$) (see Materials and Methods and Supplementary Material, Fig. S1). A total of 2.5 million single nucleotide polymorphisms (SNPs) were analyzed, most through imputations based on the HapMap CEU samples (10). SNPs showing suggestive association ($P < 10^{-4}$) with coffee consumption are listed in Supplementary Material, Table S1 (QQ-plot in Supplementary Material, Fig. S2). Two correlated SNPs, rs2470893-A and rs2472297-T ($D' = 1.00$, $r^2 = 0.69$ in the HapMap CEU samples (10), $D' = 0.99$ and $r^2 = 0.85$ in Iceland), at chromosome 15q24 satisfied our threshold for significance after adjusting for the number of SNPs tested ($P < 5 \cdot 10^{-8}$) (Fig. 1, Table 1, Supplementary Material, Table S1). Additionally, suggestive association was observed between increased coffee consumption and a variant, rs6968865-T, at 7p21.

We followed these associations up by genotyping rs2472297 and rs6968865 in a second Icelandic sample set ($n = 2430$) and a Danish sample set consisting of pregnant women ($n = 1620$). The association of both SNPs with increased coffee consumption was nominally significant in the combined follow-up sample sets ($P = 5.0 \cdot 10^{-5}$ for rs2472297 and $P = 1.4 \cdot 10^{-6}$ for rs6968865). Combining the genome-wide and follow-up sample sets ($n = 10\,661$), the association of both SNPs satisfied our threshold for genome-wide significant association (Table 1).

No interaction effect between either rs2472297 or rs6968865 and sex was observed on coffee consumption among the Icelandic and Dutch sample sets ($P = 0.47$ for rs2472297 and $P = 0.21$ for rs6968865,). Also, rs2472297-T and rs6968865-T did not predispose to drinking coffee, *per se*, in the combined Iceland, Dutch and Danish sample sets (n drinkers $= 7940$, n non-drinkers $= 956$, $P = 0.63$ for rs2472297 and $P = 0.85$ for rs6968865).

Cup size, coffee bean types and caffeine concentration vary between countries making effect size estimates somewhat difficult to compare. Despite these differences, the effect estimates obtained here for both SNPs are rather similar in all five populations tested, at around 0.2 cups a day per allele. Transforming the effect within each population to fractions of standard deviations of the sex-adjusted coffee consumption within that population, we estimate that the fraction of the variance of coffee consumption explained by rs2472297 and rs6968865 to be 0.55 and 0.46%, respectively, or 1.0% combined. Assuming comparable effects between study groups, a significant population heterogeneity was observed for rs6968865 ($P = 0.022$, $I^2 = 61.4\%$), but not for rs2472297 ($P = 0.58$, $I^2 = 0\%$). For rs6968865, the lowest effect was observed in the Sorbs [0.02 cups per allele, 95% confidence interval (CI): (−0.11, 0.15)] where rs6968865 was imputed from chip genotypes (info = 0.45) and the highest effect was observed in the Icelandic follow-up population [0.32 cups per allele, 95% CI: (0.20, 0.45)] where the SNP was genotyped directly.

The SNP rs2472297 is located at 15q24 between *CYP1A1* and *CYP1A2* in a gene-rich region of relatively low recombination rate (Fig. 1). Of the four phase II HapMap population samples, the T allele of rs2472297 is only present in the European CEU samples (10). None of the other SNPs in the region remained significantly associated with coffee consumption after accounting for the effect of rs2472297-T (Fig. 1). *CYP1A1* and *CYP1A2* are oriented head to head and separated by 23 kb. *CYP1A1* is involved in the metabolic activation of aromatic hydrocarbons. It is a homolog of *CYP1A2* and the two are thought to be co-regulated; however, a role for *CYP1A1* in caffeine metabolism has not been established. *CYP1A2* is a major hepatic enzyme that accounts for 8–15% of the total P450 content (11), is the main caffeine metabolism enzyme and is involved in the metabolism of widely used drugs (12).

The SNP rs6968865 is located within a linkage disequilibrium (LD) block of 240 kb on 7p15 (Fig. 1). *AHR*, the gene encoding the aryl hydrocarbon receptor (AHR), is the only gene located within this LD block. AHR plays a central role in xenobiotic metabolism and induces members of the CYP1 family of genes (*CYP1A1*, *CYP1A2* and *CYP1B1*) (13,14). A complex, including AHR, binds to a dioxin responsive element (DRE) resulting in transcriptional activation and the interval between *CYP1A1* and *CYP1A2* contains 13 such elements (Fig. 2) (13).

Caffeine clearance is induced by cigarette smoking and increases with the quantity of cigarettes smoked (15–17). Information on whether the individual had ever smoked cigarettes (smoking initiation) and on the number cigarettes smoked per day was available for a subset of the Icelandic ($n = 2582$) and Dutch ($n = 2875$) sample sets. Smoking initiation and coffee drinking were correlated in both populations (correlation = 0.13, $P = 4.2 \cdot 10^{-13}$ in Iceland and correlation = 0.08, $P = 1.7 \cdot 10^{-5}$ in the Netherlands) as well as the amount of coffee and cigarette consumption among individuals who both drink coffee and smoke cigarettes ($n = 1580$, correlation = 0.15, $P = 1.1 \cdot 10^{-9}$ in Iceland and $n = 2239$, correlation = 0.16, $P = 5.9 \cdot 10^{-12}$ in the Netherlands). Stratifying on smoking status, we tested the association of rs2472297-T and rs6968865-T with coffee consumption (Supplementary Material, Fig. S3). A significant association was observed in both smokers ($n = 3513$, $P = 1.9 \cdot 10^{-6}$ for rs2472297 and $P = 0.00063$ for rs6968865) and non-smokers ($n = 1209$, $P = 0.00089$ for rs2472297 and $P = 0.00046$ for rs6968865). No significant difference was observed in the association of either SNP with coffee consumption between smokers and non-smokers ($P = 0.83$ for rs2472297-T and $P = 0.66$ for rs6968865-T).
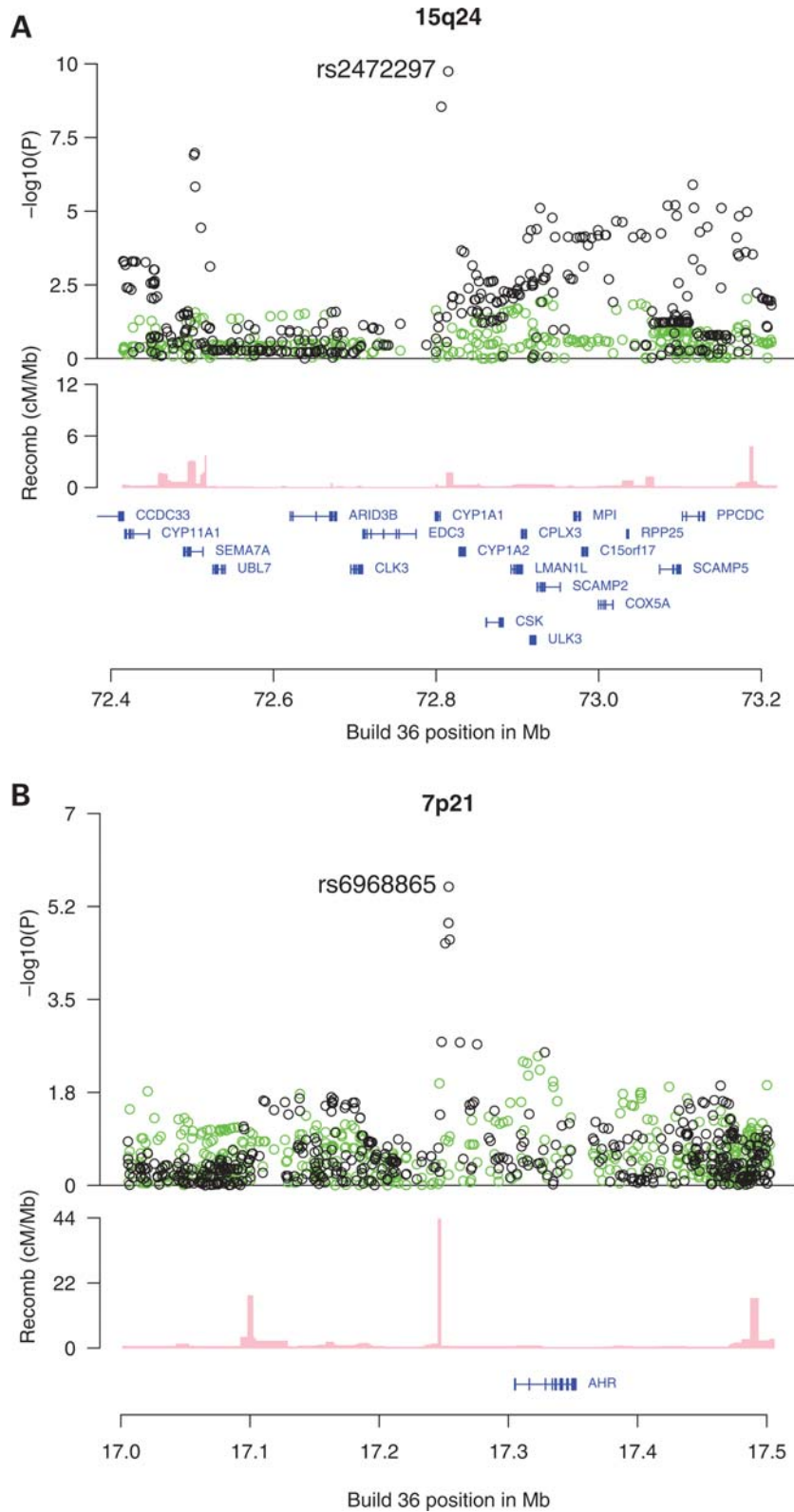
Figure 1. The genomic regions of association with coffee consumption. (**A**) The region on 15q24 around rs2472297 and (**B**) the region on 7p21 around rs6968865. Shown are the $-\log_{10}$ association *P*-values of SNPs in the region with coffee consumption from the genome-wide association meta analysis (black circles), $-\log_{10}$ association *P*-values conditional on the primary associating SNP (green circles), the SNP build 36 coordinates, the genes in the region and their exons (in blue) and recombination rates in centimorgans (cM) per megabase (Mb) (pink histogram).

Table 1. Association of rs2472297-T and rs6968865-T with coffee consumption

Population	n	Male	Age (SD)	Drink coffee	Consumption (SD)		rs2472297-T between CYP1A1 and CYP1A2				rs6968865-T near AHR			
					Male	Female	Freq	Effect (95% CI)	P-value	Info	Freq	Effect (95% CI)	P-value	Info
Discovery														
Iceland	2680	27	59 (14)	86	3.9 (2.1)	3.2 (2.0)	0.292	0.26 (0.12, 0.39)	0.00011	0.97	0.632	0.13 (0.02, 0.24)	0.025	1.00
The Netherlands	2791	62	61 (10)	97	4.2 (2.1)	3.6 (1.9)	0.265	0.31 (0.17, 0.44)	$4.3 \cdot 10^{-6}$	0.96	0.609	0.26 (0.15, 0.36)	$3.1 \cdot 10^{-6}$	1.00
Sorbs	771	40	48 (18)	85	1.9 (0.8)	1.8 (0.7)	0.158	0.24 (0.07, 0.41)	0.0046	0.39	0.593	0.02 (−0.11, 0.15)	0.77	0.45
USA	369	51	47 (7)	55	2.5 (1.8)	2.3 (1.8)	0.267	0.03 (−0.26, 0.33)	0.83	Gtd	0.607	0.12 (−0.13, 0.38)	0.34	1.00
Combined*	6611	–	–	–	–	–	–	–	$1.8 \cdot 10^{-10}$	–	–	–	$2.4 \cdot 10^{-6}$	–
Follow up														
Iceland	2430	43	50 (14)	81	4.0 (2.1)	3.3 (2.0)	0.293	0.20 (0.07, 0.33)	0.0026	Gtd	0.640	0.32 (0.20, 0.45)	$3.3 \cdot 10^{-7}$	Gtd
Denmark	1620	0	30 (4)	57	–	3.4 (2.6)	0.285	0.28 (0.08, 0.48)	0.0064	Gtd	0.665	0.14 (−0.06, 0.34)	0.17	Gtd
Combined*	4050	–	–	–	–	–	–	–	$5.0 \cdot 10^{-5}$	–	–	–	$1.4 \cdot 10^{-6}$	–
Discovery and follow up														
Combined*	10 661	–	–	–	–	–	–	–	$5.4 \cdot 10^{-14}$	–	–	–	$2.3 \cdot 10^{-11}$	–

Results are given for the four genome-wide association sample sets and two follow-up sample sets. Shown are the number of coffee drinkers (*n*), fraction of coffee drinkers that are males (male), their mean age and standard deviation (age and SD), the fraction of the individuals queried that are coffee drinkers (drink coffee), the mean and SD of the coffee consumption in cups per day for males and females (consumption and SD), the allele frequencies (freq), the estimated allelic effects on coffee consumption and its 95% CI in cups per day (effect and 95% CI), the *P*-value for the test of association (*P*-value) and the SNP genotyping imputation information (Gtd for directly genotyped SNPs).
*P-values were combined using the weighted z-scores.
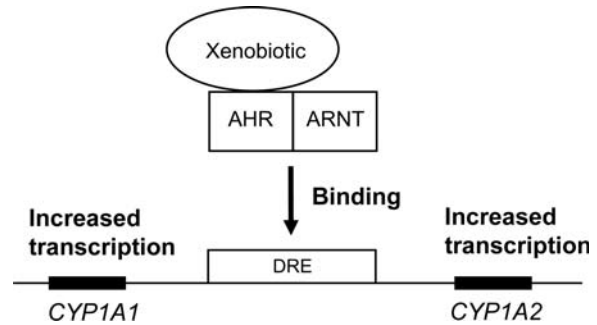
Figure 2. A schematic overview of the induction of *CYP1A1* and *CYP1A2* by the xenobiotic–AHR complex. The complex of AHR, ARNT (AHR nuclear translocator) and a xenobiotic substance, such as polycyclic aromatic hydrocarbons, binds to the DNA at the DRE and induces the transcription of *CYP1A1* and *CYP1A2*.

A sequence variant, rs1051730-A, near *CHRNA5* was previously associated with smoking quantity (18,19). We tested the association of this variant in the present genome-wide coffee consumption scan data and found nominally significant association of the allele associating with more cigarettes per day with greater coffee consumption (*P* = 0.0030). This is consistent with rs1051730-A having a role in inducing caffeine clearance through increased smoking quantity.

Two correlated SNPs in the region around *CYP1A2* were recently associated with blood pressure and hypertension: rs1378942 and rs6495122 [D′ = 0.96 and r^2 = 0.64 in the HapMap CEU samples (10)] (20,21). As these phenotypes may be influenced by coffee consumption, we postulated that these SNPs might have an effect on coffee intake. The SNP rs1378942 was not in strong LD with our coffee consumption associating SNPs (Supplementary Material, Table S2). rs1378942 was present in our genome-wide scan and its C allele, which associates with higher blood pressure and increased risk of hypertension, associates with decreased coffee consumption in our data (*P* = 0.0016). This association does not remain significant after adjusting for rs2472297 in the Icelandic and Dutch genome-wide sample sets (*P* > 0.11).

There is a large degree of inter-individual variability of CYP1A2 enzymatic activity and drug metabolism (12), and the half-life of caffeine varies substantially between individuals (3– 11 h). Determination of caffeine clearance is the most common method for CYP1A2 phenotyping (12). The variant *CYP1A2*1F* (rs762551) has been suggested to correlate with increased CYP1A2 enzymatic activity in smokers (15). *CYP1A2*1F* associates nominally with increased coffee consumption in our genome-wide scan data (*P* = 0.0080), but this association does not remain significant after adjusting for rs2472297 in the Icelandic and Dutch genome-wide sample sets (*P* = 0.21). Conversely, the association of rs2472297 with coffee consumption remained highly significant after adjusting for *CYP1A2*1F*.

DISCUSSION

We have described sequence variants associating with coffee consumption at *CYP1A1–CYP1A2* and at *AHR*, genes that encode members of the same biochemical pathway. AHR is known to induce *CYP1A1* and *CYP1A2* by binding to the DNA in the region between those two genes. Heavy coffee

consumers have higher CYP1A2 activity than those using less coffee (16,17). Low CYP1A2 activity has been associated with higher caffeine toxicity (22). It is possible that rs2472297-T and rs6968865-T allow people to consume more coffee through greater clearance of caffeine due to higher CYP1A1 or CYP1A2 enzymatic activity, but this remains to be proven experimentally.

MATERIALS AND METHODS

Iceland

The data on self-reported coffee consumption were collected from 5975 Icelanders through a questionnaire and 5110 of these classified themselves as coffee drinkers. The question asked was: 'How much coffee (with caffeine) do you usually drink?' and four answers were allowed [none/almost none; 1−7 cups or glasses a week (0.8 in regression analysis); 2−3 cups or glasses/day (2.5 in regression analysis); 4 or more cups or glasses/day (6 in regression analysis)]. The questionnaire data were collected in studies of breast cancer ($n = 4418$), skin cancers ($n = 1499$), restless leg syndrome ($n = 1280$) and ankolysing spondylitis ($n = 202$), with some individuals participating in more than one project. For individuals participating many times, the maximum reported coffee consumption was used. Out of the 5110 coffee drinkers, 2680 have been genotyped on a chip containing the Illumina 317K set of SNPs in one of several genome-wide association studies conducted by deCODE Genetics. The remaining 2430 Icelandic coffee drinkers, who had not been chip genotyped, were then genotyped for follow up using a single-track assay (Nanogen—Centaurus). The Icelandic cigarette smoking data have been described previously (19).

These studies were approved by the Data Protection Commission of Iceland and the National Bioethics Committee of Iceland. Written informed consent was obtained from all participants. Personal identifiers associated with phenotypic information and blood samples were encrypted using a third-party encryption system as previously described.

The Netherlands

Participants come from two sources previously used in genome-wide association scan for bladder cancer (23). The first set was recruited from Dutch bladder cancer patients recruited in 2007 from the population-based cancer registry held by the Comprehensive Cancer Centre East in Nijmegen. The cases were diagnosed between 1995 and 2006. Informed consent was obtained for the collection of questionnaire data on lifestyle, medical history and family history, the collection of two 10 ml blood samples, linkage to population and disease registries (cancer registry, mortality registry, hospital information systems and the Dutch demographic register). Two coffee consumption questions were answered by 1083 individuals, 1071 of which stated that they drink more than one cup of coffee a month. The questions asked were:

'How frequently, until 2 years before the diagnosis of cancer, did you consume coffee?' and five answers were allowed [never/ less than 1 cup per month; 1−3 cups per month (0.07 in regression analysis); 1 cup per week (0.14 in regression analysis); 2−4 cups per week (0.43 in regression analysis); 5−6 cups per week (0.79 in regression analysis); 1 or more cups per day]. The individuals answering '1 or more cups per day' were then asked: 'How many cups of coffee per day?' and the answer used in regression analysis. For those reporting drinking more than eight cups of coffee per day, a value of eight cups a day was used in the regression analysis.

The second set came from the Nijmegen Biomedical Study, which is a survey of the general population performed in 2002−2003 by the Radboud University Nijmegen Medical Centre. This survey was based on an age-stratified random sample of the population of Nijmegen. Genotype data were available for 1798 of these individuals, 1720 of which stated they drink coffee when asked the following two questions: 'Do you drink coffee?' and for the coffee drinkers: 'How many cups of coffee do you drink per day?', with four answers being allowed [1−2 cups (1.5 in regression analysis); 3−5 cups (4 in regression analysis); 6−8 cups (7 in regression analysis); 8 or more cups (8 in regression analysis)].

Similar informed consent as described above was obtained from these individuals and all individuals were fully informed about the goals and the procedures of the studies. All study protocols were approved by the Institutional Review Board of the Radboud University Nijmegen Medical Centre.

The Dutch cigarette smoking data have been described previously (19).

The USA

Subjects consisted of families identified between 1983 and 2006 from probands with a coronary disease before 60 years of age. This ongoing prospective family study was designed to determine the environmental and genetic causes of premature chronic and cardiovascular diseases. Probands with documented coronary artery disease (CAD) before the age of 60 were identified at the time of hospitalization in any of 10 Baltimore area hospitals. Their apparently healthy 30- to 59-year-old siblings without known CAD were recruited. In 2002, adult offspring over 21 years of age of all participating siblings and probands were recruited and underwent risk factor measurement and phenotypic characterization. In addition, the spouse was recruited for all participants for which at least one offspring was recruited.

The coffee consumption was assessed through Block Health Habits and History Questionnaire or Block Food Questionnaire.

Between 1991 and 1996, the data were collected using the Block Health Habits and History Questionnaire using the following questions on regular (caffeinated) coffee: 'How often do you usually drink Coffee, not decaffeinated?', 'Check whether your usual serving size is small, medium or large'. Between 1998 and 2003, the data were collected using the Block Food Questionnaire using the following questions on regular (caffeinated) coffee: 'How often do you drink?' with nine allowed answers (never few/year; once/month; 2−3 times/month; once/week; twice/week; 3−4 times/week; 5−6 times/week; every day) and for the individuals drinking coffee every day four answers were allowed (1 cup; 2 cups; 3−4 cups; 5+ cups). The data were available for 565 individuals of European descent, of which 369 drink coffee at least one cup of coffee a month.

Sorbs

All subjects are part of a sample from an extensively pheno-typed isolated population from Eastern Germany, the Sorbs. The Sorbs are of Slavonic origin, and have lived in ethnic isolation among the Germanic majority during the past 1100 years. Today, the Sorbian speaking, Catholic minority comprises ∼15 000 full-blooded Sorbs resident in about 10 villages in rural Upper Lusatia (Oberlausitz), Eastern Saxony. At present, >1000 Sorbian individuals are enrolled in the study. Nine hundred and thirteen subjects (540 females and 373 males) had available phenotypic information for the present analysis. Coffee consumption was assessed by two questions: 'Do you drink coffee regularly? (Yes/ No)' and 'If yes, how many cups per day on average?' Four answers were allowed: 1 cup cups (1 in regression analysis); 2 cups (2 in regression analysis); 3–4 cups (3.5 in regression analysis); more than 4 cups (6 in regression analysis). Information on quantitative coffee consumption was available for 771 individuals. The study was approved by the ethics committee of the University of Leipzig, and all subjects gave written informed consent before taking part in the study.

Denmark

The Danish National Birth Cohort (DNBC) is a population-based cohort of 101 042 pregnancies, recruited in the years 1996–2002. All participating women underwent thorough phenotype characterization based on information from four computer-assisted telephone interviews conducted during pregnancy and after delivery. The data were available for 2841 women, 1620 of which stated that they drink coffee. Their coffee consumption was ascertained in an interview with the question 'How many cups of coffee do you drink per day?' asked at the 12th and 30th weeks of pregnancy, and the maximum reported coffee consumption was used. For women reporting drinking more than five cups of coffee per day, a value of five cups a day was used in the regression analysis.

Genome-wide genotyping

The Icelandic and Dutch were assayed with the Illumina HumanHap300 or HumanHapCNV370 bead chips (Illumina, SanDiego, CA, USA). The US samples were genotyped on the Illumina Human1M chip. SNPs were excluded if they had (a) yield <95%, (b) minor allele frequency <1% in the population, or (c) showed significant deviation from Hardy–Weinberg equilibrium in the controls ($P < 0.001$). Any samples with a call rate below 98% were excluded from the analysis.

The Sorbs were assayed with the 500K Affymetrix Gene-Chip (250K Sty and 250K Nsp arrays, Affymetrix, Inc, $n = 526$) and Affymetrix Genome-Wide Human SNP Array 6.0 ($n = 494$) at the Microarray Core Facility of the Interdisciplinary Centre for Clinical Research, University of Leipzig, Germany and ATLAS Biolabs GmbH, Berlin, Germany. Genotypes were called using BRLMM algorithm (Affymetrix, Inc) for 500K and Birdseed Algorithm for Genome-Wide Human SNP Array 6.0. The quality control criteria for the genotyping data were as follows: sample call rate >94%, ethnic outliers, duplicates, gender mismatch; 941 subjects passing the genotyping QC criteria were used.

Single SNP genotyping

Single SNP genotyping for all samples was carried out at deCODE Genetics in Reykjavik, Iceland, applying the same platform to all populations studied. All single SNP genotyping was carried out using the Centaurus (Nanogen) platform (24). The quality of each Centaurus SNP assay was evaluated by genotyping each assay on the CEU samples and comparing the results with the HapMap data (10). All assays had mismatch rate <0.5%. Additionally, all markers were re-genotyped on >10% of samples typed with the Illumina platform resulting in an observed mismatch in <0.5% of samples.

SNP imputation

SNP imputation was based on the Phase II CEU HapMap samples (10). Genotypes were imputed for a total of 2.5 million SNPs using the IMPUTE software (25).

Association analysis

For the quantitative trait association analysis, i.e. coffee consumption measured in cups per day and smoking quantity measured in cigarettes per day, a classical linear regression, using the genotype as an additive covariate (or expected allele count for imputed SNPs) and the coffee consumption in cups per day as a response, was fit to test for association. An additive model for SNP effects was assumed in all instances. All associations with quantitative traits were performed adjusting for sex and age. The association analysis was performed using SNPTEST in the Sorbs data set (25). Regression analysis was also performed using the expected allele count of rs2472297-T as a covariate, in order to test for additional associations with coffee consumption.

Since the scale of coffee consumption is not uniform across populations, combined significance levels were calculated by weighing z-scores by the inverse of the square root of each study's effective sample size.

For case control association analysis, i.e. drinking versus not drinking coffee, we utilized a standard likelihood ratio statistic, implemented in the NEMO software (26) to calculate two-sided *P*-values for each individual allele, assuming a multiplicative model for risk, i.e. that the risk of the two alleles a person carries multiplies (27).

Correction for relatedness of the subjects and genomic control

We estimated an inflation factor for each genome-wide association scan by calculating the average of the 2 528 522 chi-square statistics, which is a method of genomic control (28) to adjust for both relatedness and potential population stratification. The inflation factor for coffee consumption was estimated as 1.05, 1.01, 1.04 and 1.02 in Iceland, the Netherlands, the USA and Germany, respectively, and all the results presented from association with these traits were adjusted based on these inflation factors.

Fraction of variance explained

The fraction of variance explained by each SNP was estimated using the formula: $2 f (1 - f) \beta^2$, where f is the frequency of

the effect allele of the SNP and β is the effect as a fraction of the standard deviation of coffee consumption. The minor allele frequency of each SNP was estimated by taking the mean of the estimated allele frequencies. The effect in each population was estimated by standardizing the coffee consumption within each population and sex to have standard deviation 1 and averaging the effects using the inverse of the estimated variances of the effect estimate within each population as weights.

Heterogeneity calculations

Heterogeneity was tested by comparing the null hypothesis of the effect being the same in all populations to the alternative hypothesis of each population having a different effect using a likelihood ratio test. I^2 lies between 0 and 100% and describes the proportion of total variation in study estimates that is due to heterogeneity (29).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

Conflict of Interest statement. None declared.

FUNDING

REFERENCES

1. van Dam, R.M. (2008) Coffee consumption and risk of type 2 diabetes, cardiovascular diseases, and cancer. *Appl. Physiol. Nutr. Metab.*, **33**, 1269–1283.
2. Snyder, S.H., Katims, J.J., Annau, Z., Bruns, R.F. and Daly, J.W. (1981) Adenosine receptors and behavioral actions of methylxanthines. *Proc. Natl Acad. Sci. USA*, **78**, 3260–3264.
3. Cnattingius, S., Signorello, L.B., Anneren, G., Clausson, B., Ekbom, A., Ljunger, E., Blot, W.J., McLaughlin, J.K., Petersson, G., Rane, A. *et al.* (2000) Caffeine intake and the risk of first-trimester spontaneous abortion. *N. Engl. J. Med.*, **343**, 1839–1845.
4. Riksen, N.P., Rongen, G.A. and Smits, P. (2009) Acute and long-term cardiovascular effects of coffee: implications for coronary heart disease. *Pharmacol. Ther.*, **121**, 185–191.
5. Rosso, A., Mossey, J. and Lippa, C.F. (2008) Caffeine: neuroprotective functions in cognition and Alzheimer's disease. *Am. J. Alzheimers Dis. Other Demen.*, **23**, 417–422.
6. Kawachi, I., Willett, W.C., Colditz, G.A., Stampfer, M.J. and Speizer, F.E. (1996) A prospective study of coffee drinking and suicide in women. *Arch. Intern. Med.*, **156**, 521–525.
7. Carmelli, D., Swan, G.E., Robinette, D. and Fabsitz, R.R. (1990) Heritability of substance use in the NAS-NRC Twin Registry. *Acta Genet. Med. Gemellol. (Roma)*, **39**, 91–98.
8. Laitala, V.S., Kaprio, J. and Silventoinen, K. (2008) Genetics of coffee consumption and its stability. *Addiction*, **103**, 2054–2061.
9. Vink, J.M., Staphorsius, A.S. and Boomsma, D.I. (2009) A genetic analysis of coffee consumption in a sample of Dutch twins. *Twin Res. Hum. Genet.*, **12**, 127–131.
10. (2005) A haplotype map of the human genome. *Nature*, **437**, 1299–1320.
11. Rodrigues, A.D. (1999) Integrated cytochrome P450 reaction phenotyping: attempting to bridge the gap between cDNA-expressed cytochromes P450 and native human liver microsomes. *Biochem. Pharmacol.*, **57**, 465–480.
12. Faber, M.S., Jetter, A. and Fuhr, U. (2005) Assessment of CYP1A2 activity in clinical practice: why, how, and when? *Basic Clin. Pharmacol. Toxicol.*, **97**, 125–134.
13. Nukaya, M. and Bradfield, C.A. (2009) Conserved genomic structure of the Cyp1a1 and Cyp1a2 loci and their dioxin responsive elements cluster. *Biochem. Pharmacol.*, **77**, 654–659.
14. Nukaya, M., Moran, S. and Bradfield, C.A. (2009) The role of the dioxin-responsive element cluster between the Cyp1a1 and Cyp1a2 loci in aryl hydrocarbon receptor biology. *Proc. Natl Acad. Sci. USA*, **106**, 4923–4928.
15. Sachse, C., Brockmoller, J., Bauer, S. and Roots, I. (1999) Functional significance of a C->A polymorphism in intron 1 of the cytochrome P450 CYP1A2 gene tested with caffeine. *Br. J. Clin. Pharmacol.*, **47**, 445–449.
16. Djordjevic, N., Ghotbi, R., Bertilsson, L., Jankovic, S. and Aklillu, E. (2008) Induction of CYP1A2 by heavy coffee consumption in Serbs and Swedes. *Eur. J. Clin. Pharmacol.*, **64**, 381–385.
17. Tantcheva-Poor, I., Zaigler, M., Rietbrock, S. and Fuhr, U. (1999) Estimation of cytochrome P-450 CYP1A2 activity in 863 healthy Caucasians using a saliva-based caffeine test. *Pharmacogenetics*, **9**, 131–144.
18. Saccone, S.F., Hinrichs, A.L., Saccone, N.L., Chase, G.A., Konvicka, K., Madden, P.A., Breslau, N., Johnson, E.O., Hatsukami, D., Pomerleau, O. *et al.* (2007) Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum. Mol. Genet.*, **16**, 36–49.
19. Thorgeirsson, T.E., Geller, F., Sulem, P., Rafnar, T., Wiste, A., Magnusson, K.P., Manolescu, A., Thorleifsson, G., Stefansson, H., Ingason, A. *et al.* (2008) A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature*, **452**, 638–642.
20. Newton-Cheh, C., Johnson, T., Gateva, V., Tobin, M.D., Bochud, M., Coin, L., Najjar, S.S., Zhao, J.H., Heath, S.C., Eyheramendy, S. *et al.* (2009) Genome-wide association study identifies eight loci associated with blood pressure. *Nat. Genet.*, **41**, 666–676.
21. Levy, D., Ehret, G.B., Rice, K., Verwoert, G.C., Launer, L.J., Dehghan, A., Glazer, N.L., Morrison, A.C., Johnson, A.D., Aspelund, T. *et al.* (2009) Genome-wide association study of blood pressure and hypertension. *Nat. Genet.*, **41**, 677–687.
22. Carrillo, J.A. and Benitez, J. (1996) CYP1A2 activity, gender and smoking, as variables influencing the toxicity of caffeine. *Br. J. Clin. Pharmacol.*, **41**, 605–608.
23. Kiemeney, L.A., Thorlacius, S., Sulem, P., Geller, F., Aben, K.K., Stacey, S.N., Gudmundsson, J., Jakobsdottir, M., Bergthorsson, J.T., Sigurdsson, A. *et al.* (2008) Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat. Genet.*, **40**, 1307–1312.
24. Kutyavin, I.V., Milesi, D., Belousov, Y., Podyminogin, M., Vorobiev, A., Gorn, V., Lukhtanov, E.A., Vermeulen, N.M. and Mahoney, W. (2006) A novel endonuclease IV post-PCR genotyping system. *Nucleic Acids Res.*, **34**, e128.
25. Marchini, J., Howie, B., Myers, S., McVean, G. and Donnelly, P. (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.*, **39**, 906–913.
26. Gretarsdottir, S., Thorleifsson, G., Reynisdottir, S.T., Manolescu, A., Jonsdottir, S., Jonsdottir, T., Gudmundsdottir, T., Bjarnadottir, S.M., Einarsson, O.B., Gudjonsdottir, H.M. *et al.* (2003) The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat. Genet.*, **35**, 131–138.
27. Rice, J.A. (1995) *Mathematical Statistics and Data Analysis*. Wadsworth Inc., Belmont, CA.
28. Devlin, B. and Roeder, K. (1999) Genomic control for association studies. *Biometrics*, **55**, 997–1004.
29. Higgins, J.P. and Thompson, S.G. (2002) Quantifying heterogeneity in a meta-analysis. *Stat. Med.*, **21**, 1539–1558.