A Genetic Epidemiological Study of Behavioral Traits

Najaf Amin

A Genetic Epidemiological Study of Behavioral Traits

Najaf Amin

The work in this thesis was conducted at the department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands. This work was supported by the Erasmus MC, the Erasmus University Rotterdam, the Centre for Medical Systems Biology (CMSB), the Netherlands Consortium for Healthy Ageing (NCHA), the Netherlands Organization for Scientific Research (NWO), the Netherlands Genomic Initiative (NGI), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry of Health, Welfare and Sports, the European Commission and the Municipality of Rotterdam.

ISBN: 978-90-5335-382-0
Lay-out and printed by: Ridderprint BV, Ridderkerk

A Genetic Epidemiological Study of Behavioral Traits

Een genetisch-epidemiologisch onderzoek naar aspecten van gedrag

Proefschrift

ter verkrijging van de graad van doctor aan de Erasmus Universiteit Rotterdam op gezag van de rector magnificus

Prof. Dr. H.G.Schmidt

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
Woensdag 23 maart 2011 om 13:30 uur

door

Najaf Amin

geboren te Islamabad, Pakistan

Promotiecommissie

Promotoren:

Prof. Dr. Cornelia M. van Duijn

Prof. Dr. Ben A. Oostra

Overige leden:

Prof. Dr. Dorret I. Boomsma

Prof. Dr. Till Roenneberg

Dr. Henning W. Tiemeier

"Verily! In the creation of the heavens and the earth, and in the alternation of night and day, there are indeed signs for men of understanding"

(Aal-e-Imran; verse # 190)

To my mother,

Gul-e-Shagufta

Publications and manuscripts in this thesis

Chapter 1.2

Najaf Amin, Cornelia M. van Duijn , Yurii S. Aulchenko. *A genomic background based method for association analysis in related individuals*. PLoS One. 2007 Dec 5;2(12)

Chapter 2.1

Najaf Amin, Aulchenko YS, Marieke C. Dekker, Robert F. Ferdinand, van Spreeken A, Temmink AH, Verhulst FC, Ben A. Oostra, Cornelia M. van Duijn. *Suggestive linkage of ADHD to chromosome 18q22 in a young genetically isolated Dutch population*. Eur J Hum Genet. 2009 Jul;17(7):958-66.

Chapter 3.1

Najaf Amin, Maaike Schuur, Elena S. Gusareva, Aaron J. Isaacs, Yurii S. Aulchenko, Anatoly V. Kirichenko, Irina V. Zorkoltseva, Tatiana I. Axenovich, Ben A. Oostra, A. Cecile J.W. Janssens, Cornelia M. van Duijn. *A genome wide linkage study of individuals with high scores on NEO personality traits*. (Under review Mol Psychiatry).

Chapter 3.2

Najaf Amin, Viatcheslav Saviouk, Narelle K. Hansell, A. Cecile J.W. Janssens, Marleen H.M. de Moor, Pamela A. F. Madden, Irina V. Zorkoltseva, Brenda W. Penninx, Antonio Terracciano, Manuela Uda, Toshiko Tanaka, Tonu Esko, Anu Realo, Luigi Ferrucci, Michelle Luciano, Gail Davies, Andres Metspalu, Goncalo R. Abecasis, Ian J. Deary, Katri Raikkonen, Laura J. Bierut, Paul T. Costa, Jouke J. Hottenga, Gu Zhu, Anatoly V. Kirichenko, Aaron J. Isaacs, Yurii S. Aulchenko, Gonneke Willemsen, Andrew C. Heath, and Michele L Pergadia, Sarah E. Medland, Tatiana I. Axenovich, Eco de Geus, Grant W. Montgomery, Margaret J. Wright, Ben A. Oostra, Nicholas G. Martin, Dorret I. Boomsma, Cornelia M. van Duijn. *Refining genome-wide linkage intervals using GWAS identifies several loci influencing NEO Personality Dimensions.* (Under review Biological Psychiatry)

Chapter 3.3

Najaf Amin, A. Cecile J. W. Janssens, Narelle K. Hansell, Sarah E. Medland, Yurii S. Aulchenko, Gu Zhu, Nicholas G. Martin, Ben A. Oostra, Cornelia M. van Duijn. *Maternally transmitted SCD5 is associated with agreeableness*. (In preparation)

Chapter 4.1

Najaf Amin, Karla V. Allebrandt, Bertram Müller-Myhsok, Karin Hek, Maris Teder-Laving, Caroline Hayward, Tõnu Esko, Josine G. van Mill, Scott A. Melville, Nicole Vogelzangs, Harry Campbell, Ayse Demirkan, Zoran Dogas, Fran Borovecki, Jennifer E.Huffman, Yurii S. Aulchenko, Andre G. Uitterlinden, Eric Wichmann, Albert Hofman, Ben A. Oostra, Alan F. Wright, James F. Wilson, Peter

P. Pramstaller, Andrew A. Hicks, A. Cecile J.W. Janssens, Martha Merrow, Igor Rudan, Henning Tiemeier, Brenda W. Penninx, Andres Metspalu, Thomas Meitinger, Till Roenneberg, and Cornelia M. van Duijn. *Common genetic variant in HRNBP3 is associated with sleep latency*. (Submitted PLoS Genetics).

Chapter 4.2

Karla V. Allebrandt, Najaf Amin, Bertram Müller-Myhsok, Tõnu Esko, Maris Teder-Laving, Renata V.D.M. Azevedo, Caroline Hayward, Josine van Mill, Nicole Vogelzangs, Edward W. Green, Scott A. Melville, H.-Erich Wichmann, Ben A. Oostra, A. Cecile J.W. Janssens, Harry Campbell, James F. Wilson, Andrew A. Hicks, Peter P. Pramstaller, Zoran Dogas, Igor Rudan, Martha Merrow, Brenda Penninx, Charalambos P. Kyriacou, Andres Metspalu, Cornelia M. van Duijn, Thomas Meitinger and Till Roenneberg. *ABCC9 modulates sleep duration in humans and in Drosophila*. (Under review PNAS)

Chapter 4.3

Najaf Amin, Enda Byrne, Julie Johnson, Georgia Chenevix-Trench, Stefan Walter, Ilja M. Nolte, kConFab Investigators, Jacqueline M. Vink, Rajesh Rawal, Massimo Mangino, Alexander Teumer, Joost C. Keers, Germaine Verwoert, Sebastian Baumeister, Reiner Biffar, Astrid Petersmann, Norbert Dahmen, Angela Doering, Aaron J. Isaacs, Linda Broer, Naomi R. Wray, Grant W. Montgomery, Daniel Levy, Bruce M. Psaty, Vilmundur Gudnason, Aravinda Chakravarti, Unnur Thorsteinsdottir, Frank J.A. van Rooij, Yurii S. Aulchenko, Jouke J. Hottenga, Fernando R. Rivadeneira, Albert Hofman, Andre G. Uitterlinden, Chris J. Hammond, So-Youn Shin, Arfan Ikram, Jacqueline C.M. Witteman, A. Cecile J.W. Janssens, Harold Snieder, Henning Tiemier, Bruce H. R. Wolfenbuttel, Ben A. Oostra, Andrew C. Heath, Erich Wichmann, Tim D. Spector, Hans J. Grabe, Dorret I. Boomsma, Nicholas G. Martin, Cornelia M. van Duijn. *Genome-wide association and gene expression analyses suggest association of CYP1A1/CYP1A2 with coffee drinking.* (Submitted Mol Psychiatry).

Chapter 5

Najaf Amin, Cornelia M. van Duijn, A. Cecile J.W. Janssens. *Genetic scoring analysis: a way forward in genome wide association studies*? Eur J Epidemiol. 2009;24(10):585-7.

Table of contents

Chapter 1

Introduction

Chapter 1.1

Introduction and Methods

INTRODUCTION AND METHODS

Human behavioural genetics aims to unravel the genetic and environmental contributions to variations in human behaviour. Behaviour is a complex trait, involving multiple genes that are affected by a variety of other factors. Genetic epidemiological research of behaviour goes back to Sir Francis Galton (1822-1911), who systematically studied heredity of human behaviour and mind, introducing major statistical concepts such as correlation and regression towards the mean. After a brief period in which genetic research of behaviour fell into dismay as a result of its association to eugenics and the Nazi regime, there is at present growing scientific interest in genes and behaviour. Without a doubt, behaviour in humans and animals is for a large part genetically determined[1]. In this thesis, I have limited myself to a number of diseases and traits.

Attention Deficit Hyperactivity Disorder (ADHD)

One of the most common psychiatric disorders that causes significant behavioural and social impairment throughout the life[2], particularly in early childhood and adolescence is Attention Deficit/Hyperactivity Disorder (ADHD). ADHD is a complex disorder in which both environmental and genetic factors are implicated. ADHD is characterized by inappropriate levels of inattention, hyperactivity and impulsivity that has an onset in childhood and that has major impact on aspects of behavior including learning. The disease is recognized increasingly, particularly among boys. ADHD is affecting between 8-12 percent of children worldwide, predominantly boys and is one of the most heritable psychiatric disorders.

A large number of candidate gene studies have been conducted targeting proteins that are implicated in the pathology of ADHD, e.g. dopamine. Genes studied most are the dopamine transporter gene (*DAT1*), which maps to 5p15, the dopamine D4 (*DRD4*, 11p15), and D5 (*DRD5*, 4p16) receptor genes. These studies have yielded a number of replicated findings but meta-analyses show that the associated variants are of small effect sizes, with odds ratio ranging from 1.13 to 1.9. Genome-wide linkage studies have pointed towards several loci including 1p36, 2q21, 2q35, 4q13.2, 5p13, 5q13.1, 5q33.3, 6q12, 6q22-23, 7p13, 7q21, 9q22, 11q22, 13q12, 14q12, 15q15, 16q23, 17p11, and several other regions with nominally significant evidence of linkage. Finally, genome-wide association studies (GWAS) have been conducted as part of the Genetics Analysis Information Network (GAIN), a public-private partnership between the National Institutes of Health and the private sector with the goal of promoting genome mapping for various complex diseases. None of the GWAS conducted so far have yielded any significant association for ADHD. *CDH13* was however, replicated in three studies[3]. In this thesis, I have combined genome wide linkage analysis and association analyses to study the genetics of ADHD.

Personality

Research into human personality has always been an integral part of behavioral science and research goes back to Galton. Apart from influencing behaviour, personality traits are also important risk factors for many psychiatric and non-psychiatric disorders[4-30]. Personality traits are complex and heritable with heritability estimates ranging from 33 to 60%[31-35]. Various models for the assessment of personality have been developed and one of these is the five factor model that divides human personality into five basic, universal types including Neuroticism, Extraversion, Openness, Agreeableness and Conscientiousness[36], assessed using the NEO Inventory (NEO-FFI or NEO-PI-R)[36]. These five traits, which were developed based on a lexical hypothesis, are defined by six baseline traits each, for instance, Neuroticism is defined by anxiety, hostility, depression, self-consciousness, impulsiveness and vulnerability; Extraversion is defined by warmth, gregariousness, assertiveness, activity, excitement seeking and positive emotion; Openness by fantasy, aesthetics, feelings, action, ideas and values; Agreeableness by trust, straightforwardness, altruism, compliance, modesty and tender-mindedness; and Conscientiousness by competence, order, dutifulness, achievement-striving, self-discipline and deliberation.

The negative dimensionality of Neuroticism and its association with depression, addiction, digestive system disorders and cardiovascular problems made it an attractive trait to explore genetically especially in a candidate gene setting where the primary targets were the genes in the dopamine and serotonin pathways. Genes that have been implicated in personality through candidate gene studies include *SLC6A4* and *TPH1* for Neuroticism, *ADH4* for Extraversion and Agreeableness and *CHRM2* for Agreeableness and Conscientiousness. Also genome-wide linkage studies have performed somewhat better but targeted personality traits assessed with other instruments including Eysenck's personality questionnaire (EPQ)[37] (measures Neuroticism, Extraversion and psychotism or the Tri-dimensional Personality Questionnaire (TPQ)[38] (assesses novelty seeking, harm avoidance and reward dependence). These studies have discovered several loci for Neuroticism/harm avoidance (see the chapter 3.1 for a detailed overview), of which 1q, 8p, 11q, 12q and 18q were replicated for Neuroticism in various studies.

The advent of hypothesis free genome-wide association studies (GWAS) gave an opportunity to dig deep into the genetics of not particularly Neuroticism where some biological hypothesis in terms of plausible pathways was available, but also to explore other personality traits for which this information was almost entirely non-existent. The first GWAS[39] on these five personality traits did not yield much but the sample size was most likely the limiting factor (N = 3,972). Although the *CLOCK* gene was associated with Agreeableness findings were not replicated. Another two GWAS relatively small GWAS (N = 1,227 and N = 2,235) suggested association of *MAMDC1* and *NKAIN2* genes with Neuroticism[40-41]. Realizing the need for larger sample sizes, a large meta-

analysis including ~ 18,000 individuals from 10 cohorts, was performed which associated *RASA1* gene with Openness and *KATNAL2* with Conscientiousness. However, also the findings on this large study remain to be replicated. There may be various explanations for the limited success of GWAS. The first is that the sample sizes were not impressive in terms of the number of subjects studied, variants discovered and the non-reproducibility of the findings. Other explanations may be related to the fact that the contribution of rare variants to the make up of human personality may be more important than expected or that more complex mechanisms are involved. The complex mechanisms that remain to be explored are parent of origin effects and the presence of structural variants including copy number variants (CNVs).

A parent-of-origin effect, or genomic imprinting, results from epigenetic modification of the genome which, in turn, results in unequal transcription of parental alleles. For these imprinted genes, expression of the alleles is dependent upon the sex of the parent from which they were inherited[42]. It has been shown that standard association analysis provides suboptimal power to discover disease susceptibility variants that exhibit parental-origin-specific effects[43]. At present the incorporation of parent-of-origin effects in the association model requires a family based setting. Although development of methods like long range phasing have been developed, which might enable population based studies to incorporate parent-of-origin effects in the association models, currently these methods have not been applied because of their computational complexities. A Transmission Disequilibrium Test (TDT), which requires family trios, is frequently used to test for parent-of-origin effects.

CNVs are explained by genomic rearrangements that sometimes lead to a deletion, duplication, inversion or a translocation of certain genes which changes the number of copies of those genes. As a consequence, a person may carry less than two (deletion) or more than two (gain) copies of the gene (called CNVs) resulting in the loss or gain of a function. CNVs may either be inherited or caused by *de novo* mutation. Microarray data may be used to evaluate the evidence for CNVs. A loss or a gain in the probe intensities coupled with increased homozygosity or mosaicism of the genotype frequencies of the SNP probes points towards the presence of CNVs.

Linkage analysis is a powerful tool to detect rare variants and now that the next generation sequencing techniques are providing an opportunity to zoom into the DNA further, the time is ripe for finding the actual causal variants. In the chapter 3 of this thesis I use various contemporary gene mapping techniques to find susceptible genes (regions) that influence personality traits.

Sleep and caffeine use

Sleep is highly complex anabolic state generated by multiple brain regions and neurotransmitter systems[44], accentuating the growth and rejuvenation of the immune, nervous, skeletal and

muscular systems. Sleep is known to be regulated by two mechanisms: (1) a circadian process that regulates the sleep-wake cycle (i.e., for being awake during the day and asleep during night) and (2) a homeostatic process that regulates the duration of sleep according to the sleep pressure that accumulates during wakefulness, i.e. the longer the period of wakefulness during the day, the greater the sleep pressure and the greater will be the duration of sleep when it is permitted[44]. Sleep disturbances are associated with cognition[45], body mass index (BMI)[46], blood pressure, and cardiomyopathies[47-48]. Apart from psycho-somatic health, the morbidity and mortality associated with car crashes due sleepiness are high with short duration of sleep having the greatest negative impact on alertness[49]. Short sleep has also been linked to disease[50-51] and mortality[52]. Sleep latency (time to fall asleep) is associated with academic performance in children and adolescents[53] and also known as an important measure of well being[54]. Persistent increase in sleep latency is a major characteristic of sleep onset insomnia and delayed sleep phase syndrome[55-57]. Despite high heritability estimates (sleep duration $h^2 = 40\%$, sleep latency $h^2 = 44\%$) these traits remain genetically unexplored. Most genetic findings for sleep duration come from gene knockout studies in mice and drosophila; these include *NPAS2*, *BMALI*, *CLOCK* and *DEC2* (regulating sleep duration)[58]. A small-scale genome-wide scan and a candidate clock gene study reported weak associations of sleep duration with common gene variants of *PROK2*[59] and *CLOCK*[60]. So far no genetic variant has been reported for sleep latency.

Coffee/caffeine is a known sleep antagonist[61] yet the most widely used beverage worldwide with known health benefits[62-63]. It is a primary source of caffeine among adults and also a model for addiction[64-65]. Caffeine is known to have behavioral effects and many users experience beneficial effects on psychomotor speed, mood and alertness[66-68], while others experience negative effects such as insomnia, anxiety and dysphoria[69-70]. The genetic contribution to coffee intake has been estimated to be in the range of 39 to 56%. Most genetic studies have focused on caffeine and restricted the genes search primarily to polymorphisms in the *CYP1A2* gene[71-72], which is known to be involved in the metabolism of caffeine. Further, a polymorphism in the *ADORA2A* gene has been associated with caffeine consumption in a candidate gene study[73]. To date no large scale systematic effort has been made to uncover the genes associated with coffee intake.

Study Population(s)

The studies in the first part (chapters 2 & 3) of this thesis targeting ADHD and personality are essentially conducted primarily within the framework of the Genetic research in isolated population (GRIP) program. The study sample consisted of ~ 3,000 individuals who participated in the Erasmus Rucphen Family (ERF) study[74]. The ERF cohort was ascertained from a genetically isolated region in the southwest of the Netherlands. The study population descended from 20 related couples that lived in the isolate between 1850 and 1900; genealogical data, available from 1750, captures all individuals in a single 23 generation pedigree including more than

23,000 individuals. Pedigree members 18 years and older were invited to participate. Spouses were invited only for descendents who had children older than 18 years. The pedigree is marked by multiple consanguinity and increased inbreeding. Replication is crucial is modern genetic research. In this thesis we sought replication in the GAIN consortium (958 proband-parent trios) for ADHD, which included samples from eight countries including Belgium, Germany, Ireland, Israel, the Netherlands, Spain, Switzerland, and the United Kingdom. The study presented in chapter 3.2 also includes twin samples from Netherlands (NTR) and Australia (QIMR), which are described below.

The studies in the second part (chapter 4) of this thesis are conducted in a larger framework including the ERF and the Rotterdam study (RS-I & RS-II). The Rotterdam Study-I (RS-I) is a prospective population-based cohort study of 7,983 residents aged 55 years and older living in Ommoord, a suburb of Rotterdam, the Netherlands and The Rotterdam Study-II (RS-II) is a prospective population-based cohort study of respectively 3,011 residents aged 55 years and older. Other studies include:

- EGCUT: The Estonian Genome Center, University of Tartu (EGCUT) is a bio-bank consisting of data of 40000 individuals from a population based Estonian cohort aged 18 years and older.
- KORA: The KORA research project has evolved from the WHO MONICA study (Monitoring of Trends and Determinants of Cardiovascular Disease). The KORA study uses the samples from the KORA S4 survey, which is a population-based sample from the general population living in the region of Augsburg, Southern Germany.
- KORCULA: The Korcula Study is a family-based, cross-sectional study on the Dalmatian island of Korcula.
- NESDA: The Netherlands Study of Depression and Anxiety includes ~ 1800 unrelated cases with a current or remitted major depressive disorder and ~ 1800 healthy controls.
- Netherlands Twin Register (NTR), which was established in 1987 and contains information about Dutch twins and their families voluntarily taking part in research.
- ORCADES: The Orkney Complex Disease Study is a family-based cross sectional genetic epidemiological study in the isolated Scottish islands of Orkney. Genetic diversity is decreased compared to the mainland Scottish samples, consistent with high extent of endogamy.
- SPLIT: The SPLIT study is a population based cohort study of 416 individuals residing in the city of Split, Croatia.
- SHIP: The Study of Health in Pomerania is a longitudinal population-based cohort study in West Pomerania, a region in the northeast of Germany with a total population of 212,157 inhabitants.

- TwinsUK: The TwinsUK cohort consisted of a group of twins ascertained to study the heritability and genetics of age-related diseases (www.twinsUK.ac.uk). These unselected twins were recruited from the general population through national media campaigns in the UK and shown to be comparable to age-matched population singletons in terms of disease-related and lifestyle characteristics.
- QIMR: Consists of twins recruited from the Australian Twin Registry who were mailed a Health and Lifestyle Questionnaire between 1980 and 1982. Twins were recruited through national media campaigns.
- LifeLines: The LifeLines Cohort Study is a multi-disciplinary prospective population-based cohort study examining in a unique three-generation design the health and health-related behaviours of 165,000 persons living in the North East region of The Netherlands. All survey participants are between 18 and 90 years old at the time of enrollment. Recruitment has been going on since the end of 2006, and until January 2011 over 40,000 participants have been included.

A formal description of these cohorts is provided in the forthcoming chapters.

Statistical Analysis

From a statistical perspective various approaches can be distinguished in gene discovery. In family based studies, linkage analysis can be conducted. In this analysis we study the segregation of the trait with a genomic region that encompasses a mutation involved in the disease. The approach is most powerful to detect major mutations, that are relatively rare and have a impact on the disease risk. In population-based studies, association analyses can be conducted, in which certain variants are related to the presence of disease.

Linkage Analysis

Linkage analysis is based on the fact that alleles at the same locus on a chromosome should co-segregate at a rate (recombination rate) that is directly proportional to the distance between these alleles. The objective of linkage analysis is to estimate the recombination rate (conventionally denoted by θ). Under the null hypothesis of no linkage $\theta = 0.5$. The two alleles are said to be linked if they co-segregate more often than expected or in other words $\theta < 0.5$. Log of odds (LOD) scores are generally used to test linkage. These are estimated as

$$Z(\theta) = \log_{10}\left(\frac{L(\theta)}{L(\theta = 0.5)}\right),$$

which is simply the principal log of the likelihood ratio of observed recombination to the expected recombination events. Based on previous simulations a LOD score of greater than 3.3 is taken as evidence of significant linkage[75]. Since the estimation of recombination events requires parental information, family based setting is a prerequisite for a linkage study.

Classical linkage requires the specification of the disease model in terms of the allele frequency and the mode of inheritance (e.g. dominant, recessive). Although the approach may be powerful to find the genetic origin of disease in families affected with a monogenetic disease, the approach is not powerful for complex diseases due to the presence of multiple phenocopies (person with the same disease but with a different (genetic) origin). To overcome this problem, non-parametric methods have been developed including variance components for quantitative outcome. Non-parametric linkage analysis relies on sharing of alleles identical by descent (IBD) or identical by state (IBS) between affected relatives

$$Z = \frac{S - \mu_0}{\sigma_0},$$

where 'S' is the number of alleles shared IBD between affected relatives, 'μ_0' and 'σ_0' is the expected value and variance of S. The variance components method uses the relationship information between two relatives to decompose the covariance between these relatives into genetic variance and environmental variance

$$COV(X_i, X_j) = \begin{cases} \sigma^2_{ga} + \sigma^2_{gd} + \sigma^2_{Ga} + \sigma^2_{Gd} + \sigma^2_e & i = j \\ \pi_{ij}\sigma^2_{ga} + \delta_{ij}\sigma^2_{gd} + 2\Phi_{ij}\sigma^2_{Ga} + \Delta_{ij}\sigma^2_{Gd} & i \neq j \end{cases},$$

where '*g*' is the random effect due to major gene linked to the locus being tested, '*G*' is the random effects due to other genes at unlinked loci and '*e*' is the environmental variance. The genetic variance is further decomposed into '*a*' additive and '*d*' dominance variances. 'π_{ij}' is the proportion of alleles IBD at the major locus, 'δ_{ij}' is the probability that both alleles are IBD, 'φ_{ij}' is the kinship coefficient between the relatives *i* and *j* in a pedigree and 'Δ_{ij}' is the expected probability that the relatives share both alleles IBD. The likelihood of the data then is expressed as

$$LogL = c - \frac{1}{2}\sum_{r=1}^{R}\log[\det(V_r)] - \frac{1}{2}\sum_{r=1}^{R}(X_r - K_r)^T V_r^{-1}(X_r - K_r\beta),$$

where 'X_r' is the vector of individual trait values for the r^{th} pedigree, 'V_r' is the variance-covariance matrix, 'K_r' is the matrix of covariates for the r^{th} pedigree and 'β' is the vector of fixed effects and '*R*' is the total number of pedigrees analyzed.

Genome-wide Association Analysis

For a population based sample the association analysis is performed using simple regression analysis (linear for quantitative traits and logistic for binary traits) with the trait/phenotype as a dependent variable and the genetic polymorphism as the independent variable

$$y_i = \alpha + \beta x + \sum_j \phi_j C_j,$$

Where 'y_i' is the phenotype of interest 'x' is the genotype at a certain locus, 'C_j' is the *jth* the confounding factor, 'β' is the effect of the polymorphism and 'φ_j' is the effect of the *jth* confounding factor.

While family based samples are a pre-requisite for linkage analysis and TDT, familial relations become a nuisance in association analysis and lead to false association signals if not accounted for in the association model. The TDT can be used for association analysis in family based data. The TDT measures the over-transmission of an allele from heterozygous parents to affected offspring. TDT typically requires family trios with father, mother and child but is modified to incorporate extended families and quantitative outcomes. TDT analysis assesses linkage in the presence of association and is not confounded by population substructure as it ignores between-family variation. Apart from testing for association it also incorporates the assumption of linkage. The statistic takes the form

$$Z = \frac{S - E(S)}{\sqrt{\text{var}(S)}},$$

where '*S*' is the allele transmission count. Since the test ignores between family variation and families with homozygous parents, this results in the loss of power of the association analysis. However for specific application such as parent-of-origin testing or haplotype association analysis the TDT is powerful. In the next chapter I propose a new and more powerful test for association analysis of quantitative traits in family based samples.

REFERENCES

1. Plomin, R., Asbury, K., and Dunn, J. (2001). Why are children in the same family so different? Nonshared environment a decade later. Can J Psychiatry 46, 225-233.
2. Lantieri, F., Glessner, J.T., Hakonarson, H., Elia, J., and Devoto, M. (2010). Analysis of GWAS top hits in ADHD suggests association to two polymorphisms located in genes expressed in the cerebellum. Am J Med Genet B Neuropsychiatr Genet 153B, 1127-1133.
3. Franke, B., Neale, B.M., and Faraone, S.V. (2009). Genome-wide association studies in ADHD. Hum Genet 126, 13-50.
4. Bienvenu, O.J., Hettema, J.M., Neale, M.C., Prescott, C.A., and Kendler, K.S. (2007). Low extraversion and high neuroticism as indices of genetic and environmental risk for social phobia, agoraphobia, and animal phobia. Am J Psychiatry 164, 1714-1721.
5. Budaev, S.V. (1999). Personality and Individual Differences. Elsevier 26, 801-813.
6. Cloninger, C.R., Svrakic, D.M., and Przybeck, T.R. (1993). A psychobiological model of temperament and character. Arch Gen Psychiatry 50, 975-990.
7. Cramer, D. (1991). Type A behaviour pattern, extraversion, neuroticism and psychological distress. Br J Med Psychol 64 (Pt 1), 73-83.
8. Crowe, M., Andel, R., Pedersen, N.L., Fratiglioni, L., and Gatz, M. (2006). Personality and risk of cognitive impairment 25 years later. Psychol Aging 21, 573-580.
9. Fanous, A.H., Neale, M.C., Aggen, S.H., and Kendler, K.S. (2007). A longitudinal study of personality and major depression in a population-based sample of male twins. Psychol Med 37, 1163-1172.
10. Goodwin, R.D., Castro, M., and Kovacs, M. (2006). Major depression and allergy: does neuroticism explain the relationship? Psychosom Med 68, 94-98.
11. Goodwin, R.D., and Stein, M.B. (2003). Peptic ulcer disease and neuroticism in the United States adult population. Psychother Psychosom 72, 10-15.
12. Griffith, J.H., Frith, C.D., and Eysenck, S.B. (1980). Psychoticism and thought disorder in psychiatric patients. Br J Soc Clin Psychol 19, 65-71.
13. Kaasinen, V., Nurmi, E., Bergman, J., Eskola, O., Solin, O., Sonninen, P., and Rinne, J.O. (2001). Personality traits and brain dopaminergic function in Parkinson's disease. Proc Natl Acad Sci U S A 98, 13272-13277.
14. Kendler, K.S., Gatz, M., Gardner, C.O., and Pedersen, N.L. (2006). Personality and major depression: a Swedish longitudinal, population-based twin study. Arch Gen Psychiatry 63, 1113-1120.
15. Lichtenstein, P., Pedersen, N.L., Plomin, R., Faire, U.D., and McClearn, G.E. (1989). Type A behavior pattern, related personality traits and self-reported coronary heart disease. Person individ Diff 10, 419-426.
16. McCrae, R.R., Yang, J., Costa, P.T., Jr., Dai, X., Yao, S., Cai, T., and Gao, B. (2001). Personality profiles and the prediction of categorical personality disorders. J Pers 69, 155-174.
17. Meins, W., and Dammast, J. (2000). Do personality traits predict the occurrence of Alzheimer's disease? Int J Geriatr Psychiatry 15, 120-124.
18. Menza, M. (2000). The personality associated with Parkinson's disease. Curr Psychiatry Rep 2, 421-426.
19. Rahman, M.A., and Eysenck, S.B. (1978). Psychoticism and response to treatment in neurotic patients. Behav Res Ther 16, 183-189.
20. Robertson, D.A., Ray, J., Diamond, I., and Edwards, J.G. (1989). Personality profile and affective state of patients with inflammatory bowel disease. Gut 30, 623-626.
21. Shipley, B.A., Weiss, A., Der, G., Taylor, M.D., and Deary, I.J. (2007). Neuroticism, extraversion, and mortality in the UK Health and Lifestyle Survey: a 21-year prospective cohort study. Psychosom Med 69, 923-931.
22. Sutin, A.R., Costa, P.T., Jr., Uda, M., Ferrucci, L., Schlessinger, D., and Terracciano, A. (2010). Personality and metabolic syndrome. Age (Dordr) 32, 513-519.
23. Sutin, A.R., Scuteri, A., Lakatta, E.G., Tarasov, K.V., Ferrucci, L., Costa, P.T., Jr., Schlessinger, D., Uda, M., and Terracciano, A. (2010). Trait antagonism and the progression of arterial thickening: women with antagonistic traits have similar carotid arterial thickness as men. Hypertension 56, 617-622.

24. Sutin, A.R., Terracciano, A., Deiana, B., Naitza, S., Ferrucci, L., Uda, M., Schlessinger, D., and Costa, P.T., Jr. (2010). High neuroticism and low conscientiousness are associated with interleukin-6. Psychol Med 40, 1485-1493.

25. Sutin, A.R., Terracciano, A., Deiana, B., Uda, M., Schlessinger, D., Lakatta, E.G., and Costa, P.T., Jr. (2010). Cholesterol, triglycerides, and the Five-Factor Model of personality. Biol Psychol 84, 186-191.

26. Terracciano, A., and Costa, P.T., Jr. (2004). Smoking and the Five-Factor Model of personality. Addiction 99, 472-481.

27. Terracciano, A., Sutin, A.R., McCrae, R.R., Deiana, B., Ferrucci, L., Schlessinger, D., Uda, M., and Costa, P.T., Jr. (2009). Facets of personality linked to underweight and overweight. Psychosom Med 71, 682-689.

28. Van Os, J., and Jones, P.B. (2001). Neuroticism as a risk factor for schizophrenia. Psychol Med 31, 1129-1134.

29. Wilson, R.S., Schneider, J.A., Arnold, S.E., Bienias, J.L., and Bennett, D.A. (2007). Conscientiousness and the incidence of Alzheimer disease and mild cognitive impairment. Arch Gen Psychiatry 64, 1204-1212.

30. Winstow, D.J., Wakefield, J.A., and Goldsmith, W.M. (1990). The relationship between personality, health symptoms and disease. Person individ Diff 11, 717-723.

31. Jang, K.L., Livesley, W.J., and Vernon, P.A. (1996). Heritability of the big five personality dimensions and their facets: a twin study. J Pers 64, 577-591.

32. Jang, K.L., McCrae, R.R., Angleitner, A., Riemann, R., and Livesley, W.J. (1998). Heritability of facet-level traits in a cross-cultural twin sample: support for a hierarchical model of personality. J Pers Soc Psychol 74, 1556-1565.

33. Loehlin, J.C., McCrae, R.R., Paul T. Costa, j., and John, O.P. (1998). Heritabilities of common and measure-specific components of the big five personality factors. Journal of research in personality 32, 431-453.

34. Riemann, R., Angleitner, A., and Strelau, J. (1997). Genetic and environmental influences on personality: A study of twins reared together using the self- and peer report NEO-FFI scales. J Personal 65, 449-475.

35. Waller, N. (1999). Evaluating the structure of personality. In Personality and psychopathology, C. Cloninger, ed. (Washington, D.C., American Psychiatric Association), pp 155-197.

36. Costa, P.T., and McCrae, R.R. (1992). Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI) Professional Manual. Psychological Assessment Resources.

37. Eysenck, S.B.G., Eysenck, H.J., and Barrett, P. (1985). A revised version of the psychotism scale. Person individ Diff 6, 21-29.

38. Cloninger, C.R., Przybeck, T.R., and Svrakic, D.M. (1991). The Tridimensional Personality Questionnaire: U.S. normative data. Psychol Rep 69, 1047-1057.

39. Terracciano, A., Sanna, S., Uda, M., Deiana, B., Usala, G., Busonero, F., Maschio, A., Scally, M., Patriciu, N., Chen, W.M., et al. (2008). Genome-wide association scan for five major dimensions of personality. Mol Psychiatry.

40. Calboli, F.C., Tozzi, F., Galwey, N.W., Antoniades, A., Mooser, V., Preisig, M., Vollenweider, P., Waterworth, D., Waeber, G., Johnson, M.R., et al. (2010). A genome-wide association study of neuroticism in a population-based sample. PLoS One 5, e11504.

41. van den Oord, E.J., Kuo, P.H., Hartmann, A.M., Webb, B.T., Moller, H.J., Hettema, J.M., Giegling, I., Bukszar, J., and Rujescu, D. (2008). Genomewide association analysis followed by a replication study implicates a novel candidate gene for neuroticism. Arch Gen Psychiatry 65, 1062-1071.

42. Reik, W., and Walter, J. (2001). Genomic imprinting: parental influence on the genome. Nat Rev Genet 2, 21-32.

43. Kong, A., Steinthorsdottir, V., Masson, G., Thorleifsson, G., Sulem, P., Besenbacher, S., Jonasdottir, A., Sigurdsson, A., Kristinsson, K.T., Frigge, M.L., et al. (2009). Parental origin of sequence variants associated with complex diseases. Nature 462, 868-874.

44. Foster, R.G. (2010). A sense of time: body, clocks, sleep and health. Deutsche medizinische Wochenschrift (1946) 135, 2601-2608.

45. Stickgold, R. (2001). Toward a cognitive neuroscience of sleep. Sleep Med Rev 5, 417-421.

46. Taheri, S., and Thomas, G.N. (2008). Is sleep duration associated with obesity-where do U stand? Sleep Med Rev 12, 299-302.

47. Gangwisch, J.E. (2009). Epidemiological evidence for the links between sleep, circadian rhythms and metabolism. Obes Rev 10 Suppl 2, 37-45.

48. Gangwisch, J.E., Malaspina, D., Posner, K., Babiss, L.A., Heymsfield, S.B., Turner, J.B., Zammit, G.K., and Pickering, T.G. (2010). Insomnia and sleep duration as mediators of the relationship between depression and hypertension incidence. Am J Hypertens 23, 62-69.

49. Gillberg, M. (1995). Sleepiness and its relation to the length, content, and continuity of sleep. J Sleep Res 4, 37-40.

50. Shankar, A., Syamala, S., and Kalidindi, S. (2010). Insufficient Rest or Sleep and Its Relation to Cardiovascular Disease, Diabetes and Obesity in a National, Multiethnic Sample. PloS one 5, e14189.

51. Sabanayagam, C., and Shankar, A. (2010). Sleep duration and cardiovascular disease: results from the National Health Interview Survey. Sleep 33, 1037-1042.

52. Ikehara, S., Iso, H., Date, C., Kikuchi, S., Watanabe, Y., Wada, Y., Inaba, Y., Tamakoshi, A., and Group, J.S. (2009). Association of sleep duration with mortality from cardiovascular disease and other causes for Japanese men and women: the JACC study. Sleep 32, 295-301.

53. Touchette, E., Petit, D., Seguin, J.R., Boivin, M., Tremblay, R.E., and Montplaisir, J.Y. (2007). Associations between sleep duration patterns and behavioral/cognitive functioning at school entry. Sleep 30, 1213-1219.

54. Palmer, C.D., and Harrison, G.A. (1983). Sleep latency and lifestyle in Oxfordshire villages. Ann Hum Biol 10, 417-428.

55. Eddy, M., and Walbroehl, G.S. (1999). Insomnia. Am Fam Physician 59, 1911-1916, 1918.

56. van den Heuvel, C.J., and Lushington, K. (2002). Chronobiology and insomnia: pathophysiology and treatment of circadian rhythm sleep disorders. Expert Rev Neurother 2, 249-260.

57. Weitzman, E.D., Czeisler, C.A., Coleman, R.M., Spielman, A.J., Zimmerman, J.C., Dement, W., Richardson, G., and Pollak, C.P. (1981). Delayed sleep phase syndrome. A chronobiological disorder with sleep-onset insomnia. Arch Gen Psychiatry 38, 737-746.

58. Hor, H., and Tafti, M. (2009). Physiology. How much sleep do we need? Science (New York, NY 325, 825-826.

59. Gottlieb, D.J., O'Connor, G.T., and Wilk, J.B. (2007). Genome-wide association of sleep and circadian phenotypes. BMC medical genetics 8 Suppl 1, S9.

60. Allebrandt, K.V., Teder-Laving, M., Akyol, M., Pichler, I., Muller-Myhsok, B., Pramstaller, P., Merrow, M., Meitinger, T., Metspalu, A., and Roenneberg, T. (2010). CLOCK gene variants associate with sleep duration in two independent populations. Biol Psychiatry 67, 1040-1047.

61. Lorist, M.M., and Snel, J. (2008). Caffeine, Sleep, and Quality of Life. In Sleep and Quality of Life in Clinical Medicine. (NJ, Humana Press), pp 325-331.

62. Choi, H.K., and Curhan, G. (2010). Coffee consumption and risk of incident gout in women: the Nurses' Health Study. Am J Clin Nutr 92, 922-927.

63. Ruusunen, A., Lehto, S.M., Tolmunen, T., Mursu, J., Kaplan, G.A., and Voutilainen, S. (2010). 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 13, 1215-1220.

64. Hughes, J.R. (1992). Clinical importance of caffeine withdrawal. N Engl J Med 327, 1160-1161.

65. Hughes, J.R., Oliveto, A.H., Helzer, J.E., Higgins, S.T., and Bickel, W.K. (1992). Should caffeine abuse, dependence, or withdrawal be added to DSM-IV and ICD-10? Am J Psychiatry 149, 33-40.

66. Horne, J.A., and Reyner, L.A. (1996). Counteracting driver sleepiness: effects of napping, caffeine, and placebo. Psychophysiology 33, 306-309.

67. Mitchell, P.J., and Redman, J.R. (1992). Effects of caffeine, time of day and user history on study-related performance. Psychopharmacology (Berl) 109, 121-126.

68. Silverman, K., Mumford, G.K., and Griffiths, R.R. (1994). Enhancing caffeine reinforcement by behavioral requirements following drug ingestion. Psychopharmacology (Berl) 114, 424-432.

69. Evans, S.M., and Griffiths, R.R. (1991). Dose-related caffeine discrimination in normal volunteers: individual differences in subjective effects and self-reported cues. Behav Pharmacol 2, 345-356.

70. Childs, E., Hohoff, C., Deckert, J., Xu, K., Badner, J., and de Wit, H. (2008). Association between ADORA2A and DRD2 polymorphisms and caffeine-induced anxiety. Neuropsychopharmacology 33, 2791-2800.

71. Cornelis, M.C., El-Sohemy, A., Kabagambe, E.K., and Campos, H. (2006). Coffee, CYP1A2 genotype, and risk of myocardial infarction. JAMA 295, 1135-1141.

72. Castorena-Torres, F., Mendoza-Cantu, A., de Leon, M.B., Cisneros, B., Zapata-Perez, O., Lopez-Carrillo, L., Salinas, J.E., and Albores, A. (2005). CYP1A2 phenotype and genotype in a population from the Carboniferous Region of Coahuila, Mexico. Toxicol Lett 156, 331-339.

73. Cornelis, M.C., El-Sohemy, A., and Campos, H. (2007). Genetic polymorphism of the adenosine A2A receptor is associated with habitual caffeine consumption. Am J Clin Nutr 86, 240-244.

74. Pardo, L.M., MacKay, I., Oostra, B., van Duijn, C.M., and Aulchenko, Y.S. (2005). The effect of genetic drift in a young genetically isolated population. Ann Hum Genet 69, 288-295.

75. Lander, E., and Kruglyak, L. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11, 241-247.

Chapter 1.2

A genomic background based method for association analysis in related individuals

Najaf Amin, Cornelia M. van Duijn and Yurii S. Aulchenko*
Department of Epidemiology & Biostatistics, Erasmus MC Rotterdam, The Netherlands.

ABSTRACT

Feasibility of genotyping of hundreds and thousands of single nucleotide polymorphisms (SNPs) in thousands of study subjects have triggered the need for fast, powerful, and reliable methods for genome-wide association analysis. Here we consider a situation when study participants are genetically related (e.g. due to systematic sampling of families or because a study was performed in a genetically isolated population). Of the available methods that account for relatedness, the Measured Genotype (MG) approach is considered the 'gold standard'. However, MG is not efficient with respect to time taken for the analysis of genome-wide data. In this context we proposed a fast two-step method called Genome-wide Association using Mixed Model and Regression (GRAMMAR) for the analysis of pedigree-based quantitative traits. This method certainly overcomes the drawback of time limitation of the measured genotype (MG) approach, but pays in power. One of the major drawbacks of both MG and GRAMMAR, is that they crucially depend on the availability of complete and correct pedigree data, which is rarely available. In this study we first explore type 1 error and relative power of MG, GRAMMAR, and Genomic Control (GC) approaches for genetic association analysis. Secondly, we propose an extension to GRAMMAR i.e. GRAMMAR-GC. Finally, we propose application of GRAMMAR-GC using the kinship matrix estimated through genomic marker data, instead of (possibly missing and/or incorrect) genealogy. Through simulations we show that MG approach maintains high power across a range of heritabilities and possible pedigree structures, and always outperforms other contemporary methods. We also show that the power of our proposed GRAMMAR-GC approaches to that of the 'gold standard' MG for all models and pedigrees studied. We show that this method is both feasible and powerful and has correct type 1 error in the context of genome-wide association analysis in related individuals.

INTRODUCTION

Most of the complex genetic diseases have risk factors that are quantitative in nature. For instance, cholesterol level is a risk factor for myocardial infarction, Body mass index is a risk factor for type 2 diabetes. These quantitative traits (QTs) often have strong genetic determinants. It is therefore, of considerable interest to map the genes underlying QTs[1]. For most QTs relevant for human health and disease a large proportion – ranging from 30 to 80% – of variation is explained by genetic factors. Multiple genes are expected to contribute to this variation. The proportion of variation explained by a single gene is expected, however, to be small (less than 5%). For example, one of most known quantitative trait loci (QTLs), APOE is strongly and consistently associated with increased total cholesterol levels. Yet it explains only about 2-5% of the variation of this trait[2-3]. When effects of specific common variants are expected to be small, association analysis provides a powerful approach to identify the gene compared to linkage analysis[4]. Genome-wide association analysis is a powerful tool to disentangle the complexity of quantitative traits, even in family based studies.

For pedigree-based association analysis several methods and software packages have been developed that utilize information about transmission of alleles, such as the orthogonal test for within-family variation [5] and family-based association test [6-7]. As these methods analyze within-family variation, they are robust to population stratification. However, these methods ignore a large proportion of information provided by the between-family variation leaving room for improvement.

A conventional polygenic model of inheritance, which is a statistical genetics'"gold standard", is a mixed model

$$y = \mu + G + e,$$

where μ is the overall mean, *G* is the vector of random polygenic effect, and *e* is the vector of random residuals. This model may be extended to study association by including a *kg* term, where *k* is the marker genotype effect, and *g* is the vector of genotypes. Such a model is known as the measured genotype (MG) model[8]. The MG approach, implemented using (restricted) maximum likelihood, is a powerful tool for the analysis of QTs when ethnic stratification can be ignored[9-10] and pedigrees are small or when there are few dozens or hundreds of candidate polymorphisms to be tested. This approach, however, is not efficient in terms of computation time. This hampers the application of MG in genome-wide association analysis.

We proposed a fast two-step approximation to MG, a Genome Wide Rapid Association using Mixed Model and Regression (GRAMMAR)[11] . In the first step the individual environmental residuals are estimated, using additive polygenic model. Then the test of association is performed using these familial correlation-free residuals with a rapid least squares or score method. Though the two-step method is indeed computationally efficient and outperforms family based approaches like FBAT and QTDT in terms of power and speed, it loses power compared to MG[11].

The test becomes increasingly conservative and less powerful with the increase in the number of large full-sib families and increased heritability of the trait. Interestingly, empirical power of GRAMMAR is very close to that of MG.

Both classical MG and GRAMMAR approaches rely heavily on the availability, completeness, and correctness of genealogical information. When these assumptions are violated, the most likely outcome is inflation in type 1 error. Practically, genealogical information may often be available only for a limited number of generations, it may be inaccurate, and it may become increasingly inaccurate back in time. This may be taken as an argument for application of underpowered TDT-like methods to avoid false positive or negative results.

We and others reason that genomic data can be used to correct for the (only partly observed) true genealogy. With the new array technologies, large numbers of markers can be typed over the genome. These provide information on 'genomic background', which can be used to infer population (sub) structure and relations between study participants, which is classically done using the genealogy. In a recent study of type 2 diabetes, genomic control (GC) was applied to control for relatedness among cases and controls from Icelandic population[12], However, the type 1 and 2 errors of GC were not yet systematically investigated in the context of pedigree data analysis. In this study we aim to exploit the ideas of genomic background to extend our work on family based association[11] and determine how powerful and efficient the method for genome-wide association analysis of QTs in samples of related individuals is.

METHODS

Only minor proportion of markers in a genome-wide association study is expected to be truly associated with an analysis trait, and a vast majority of the markers may be thought of as realizations under the null hypothesis and can be used to characterize the null distribution of the test statistics. This idea follows that of Genomic Control (GC) method[13], which was introduced in the context of association analysis in population-based studies, where population stratification and cryptic relatedness may be present.

Following Devlin and Roeder[13-14] we suggest estimation of the test statistic inflation factor λ by regressing the trait on N loci. From each regression analysis, estimate $T_i^2=\beta_i^2/Var(\beta_i)$, where β_i is the effect of the i-th SNP (i from 1 to N) and $Var(\beta_i)$ is the variance of the estimate. Inflation factor is estimated as

$$\lambda = Median\ (T_1^2, T_2^2, \ldots, T_N^2)/0.456$$

where 0.456 is the median of the χ_1^2 distribution with a non-central variance equal to φ. The number of loci used, N, in a genome-wide association study is typically reflecting all loci investigated or the ones generating lowest 95% of test statistics[15-16]. It should be noted that the value of λ in conventional GC is constrained to be greater than one as values less than one have no theoretical meaning.

We propose use of GC to correct for conservativeness of the GRAMMAR approach outlined earlier. This method which we call GRAMMAR-GC involves three steps: (a) trait heritability is estimated by using trait and pedigree data using the following mixed model

$$y_i = \mu + \sum_j \beta_j c_{ij} + G_i + e_i$$

where β_j is the effect of j^{th} covariate , c_{ij} is the value of j^{th} covariate and μ, G, and e are defined earlier. And environmental residuals are estimated as

$$y_i^* = y - (\hat{\mu} + \sum_j \beta_j c_{ij} + \hat{G}_i)$$

(b) the markers are tested for association with these residuals using simple linear regression

$$y_i^* = \mu + kg_i + e_i$$

where k and g are defined earlier

(c) GC procedure is applied to correct the test statistic. The deflation factor ζ is estimated by regressing residuals from step (a) on each of the k null loci and from each regression analysis $T_k^2 = \hat{\beta}_k^2 / var(\hat{\beta}_k)$ is estimated, where $\hat{\beta}_k$ is the effect of the k^{th} SNP. The deflation factor ζ is estimated as $\hat{\zeta} = median\ (T_1^2, T_2^2, \ldots, T_K^2)/0.456$. Then $T^2/\hat{\zeta}$ is compared with $\chi^2_{(1)}$ to determine whether the locus is significantly associated with the QT [14].

Steps (a) and (b) comprise the original GRAMMAR approach, leading to a conservative test. In step (c), GC is used to estimate the deflation factor ζ. This deflation factor is estimated in exactly the same way Bacanu et al. [14] estimate inflation factor λ for quantitative traits. The only difference is that $\zeta < 1$ in contrast to λ in conventional GC methods which is constrained to be > 1. This difference comes from the fact that residuals from step (a) are regressed on null loci to obtain the estimate of ζ instead of original trait data as in conventional GC.

The original GRAMMAR relies on the availability of a precise and complete pedigree structure for heritablity estimation in step (a). This can, however, be done by using kinship coefficients estimated from genomic data where the genomic estimate of kinship for a pair of individuals i and j is obtained using the formula[17]

$$f_{ij} = \frac{1}{n}\sum_{k=1}^{n} \frac{(g_{ik} - p_k)(g_{jk} - p_k)}{p_k(1 - p_k)}$$

where g_{ik} is the genotype of the i-th person at the k-th SNP of the (coded as 0, 1/2 and 1, for rare allele homozygote, heterozygote and common homozygote, respectively), p_k is the frequency of the major allele, and n is the number of SNPs used for kinship estimation.

Heritability is then estimated by maximizing the log-likelihood of the data

$$\log L = -\frac{1}{2}\left(\log_e|\Sigma| + (y - \mu)^T \Sigma^{-1} (y - \mu)\right)$$

where y is the vector of trait values, μ is the mean and $\Sigma = \varphi\sigma_G^2 + I\sigma_e^2$ is the variance-covariance

matrix. Here, φ is the relationship matrix whose elements are defined as 2 f_{ij}, σ_G^2 is the additive genetic variance due to polygenes, I is the identity matrix and σ_e^2 is the residual variance. Trait environmental residuals were obtained as

$$y^* = \hat{\sigma}_e^2 \, \hat{\Sigma}^{-1} \, (y - \hat{\mu})$$

To avoid confusion we refer to the new method as Pedigree GRAMMAR-GC (PedGR-GC) when in step (a) heritability is estimated from the genealogy, and Genomic GRAMMAR-GC (GenGR-GC) when the heritability is estimated from the genomic data. For PedGR-GC, environmental residuals were estimated using ASReml [18]. All other computations were performed using freely available R software (http://www.r-project.org); computations associated with GenGR-GC were facilitated by GenABEL package for R[19], implementing procedures to compute genomic kinship, maximize polygenic models and compute the residuals.

Pedigrees used

To investigate type 1 error and power of the proposed methods we used three different pedigree structures representing three different study scenarios. For example, Nuclear pedigrees (NP) simulated a study performed in the outbred population, the Erasmus Rucphen Family study (ERF) population is a study of a genetically isolated population and Idealized Pig Population (IPP) simulates a livestock population.

NP: 337 sib-trios each having 3 phenotyped and genotyped siblings; in total, 1011 individuals were available for the analysis. The founders in each pedigree were assumed to be unrelated.

ERF: 1010 phenotyped and genotyped individuals all related to each other in a single large complex pedigree of about 10,000 individuals. The pedigree extends up to 23 generations and contains thousands of loops. The phenotyped individuals are a part of Erasmus Rucphen Family (ERF) study, performed in a young genetically isolated Dutch population[20].

IPP: idealized pig population, consisting of 10 sires, each mated with 10 dams, nine of which have 10 and one 11 measured offspring. Thus each sire has 101 half-sib offspring in families of 10 full-sibs. In total 1010 phenoytped individuals were available for the analysis.

Genetic data was simulated using each of these pedigrees under several models. The SNP that was analyzed for association had a minor allele frequency of 10%. For studying type 1 error this SNP was not associated with the trait while for studying power this SNP explained 1, 2, or 3% of the total trait variation and acted in an additive manner. The total heritability of the trait was set to 30, 50, and 80% and this included the variation due to the QTL studied. To enable genomic control, we also simulated 200 unlinked SNPs for each realization.

These pedigrees and models were used to evaluate original GRAMMAR and thus we can directly compare type 1 error and power of the suggested methods to these evaluated by Aulchenko et al [11], namely, MG, DFS (linear regression which does not take family structure into account), QTDT and FBAT.

Table1: 95th percentiles of the distribution and type 1 error for PedGR-GC, GC and GRAMMAR

Pedigree			95th percentile			Type 1 error at a given threshold $\chi^2 > 3.84$			$\chi^2 > 6.64$				
	q_{snp}	h^2	PedGR-GC	GC	GRAMMAR	PedGR-GC	GC	GRAMMAR	PedGR-GC	GC	GRAMMAR	λ ± SE(λ) GC	ζ ± SE(ζ) PedGR-GC
NP													
	0.1	0.3	3.97	3.90	3.25	0.053	0.051	0.029	0.014	0.013	0.010	1.16±0.006	0.88±0.004
	0.1	0.3q	3.89	3.71	3.11	0.050	0.045	0.034	0.012	0.010	0.004	1.16±0.006	0.88±0.005
	0.1	0.5	3.61	3.70	2.89	0.040	0.048	0.021	0.004	0.006	0.002	1.26±0.007	0.82±0.004
	0.1	0.8	3.85	4.17	2.92	0.050	0.056	0.026	0.017	0.016	0.006	1.41±0.007	0.72±0.004
	0.5	0.3	3.92	3.87	3.35	0.050	0.052	0.039	0.010	0.014	0.003	1.17±0.006	0.88±0.005
	0.5	0.3q	3.93	3.73	3.14	0.051	0.048	0.033	0.011	0.009	0.005	1.15±0.006	0.88±0.005
	0.5	0.5	3.65	4.01	2.80	0.045	0.056	0.027	0.008	0.009	0.003	1.26±0.007	0.81±0.004
	0.5	0.8	3.89	4.12	2.7	0.052	0.059	0.024	0.011	0.008	0.003	1.42±0.007	0.72±0.004
ERF													
	0.1	0.3	3.85	3.78	3.08	0.050	0.047	0.031	0.009	0.010	0.004	1.29±0.008	0.86±0.005
	0.1	0.3q	3.93	3.83	3.09	0.053	0.049	0.034	0.006	0.004	0.002	1.27±0.008	0.85±0.005
	0.1	0.5	3.93	3.86	3.15	0.055	0.050	0.023	0.010	0.009	0.003	1.47±0.009	0.79±0.004
	0.1	0.8	3.95	3.90	2.71	0.052	0.051	0.021	0.012	0.009	0.001	1.71±0.010	0.69±0.003
IPP													
	0.1	0.3	4.08	3.83	2.73	0.056	0.049	0.019	0.017	0.008	0.003	3.26±0.037	0.68±0.003
	0.1	0.3q	3.66	3.06	2.43	0.043	0.026	0.014	0.011	0.007	0.001	3.19±0.046	0.68±0.004
	0.1	0.5	4.14	3.70	2.58	0.056	0.046	0.012	0.011	0.012	0.004	4.72±0.053	0.63±0.003
	0.1	0.8	4.13	3.75	2.27	0.059	0.045	0.012	0.014	0.009	0.000	7.03±0.077	0.58±0.003

Pedigree studied NP: 337 nuclear families; ERF:1010 in one large pedigree; IPP: idealized pig population
h^2: total heritability; total heritability of 0.3q represents heritability of 0.3 explained by a single unlinked QTL
q_{snp}: minor allele frequency of the SNPs studied
λ: estimate of the inflation factor for genomic control
ζ: estimate of the deflation factor for GRAMMAR-GC

RESULTS

Table 1 shows the 95th percentile of the test statistic and type 1 error (proportion of simulations that resulted in a $\chi^2 \geq 3.84$ and $\chi^2 \geq 6.64$, corresponding to tabulated $\alpha = 0.05$ and 0.01 respectively) for GC and PedGR-GC and GRAMMAR. For GC and PedGR-GC type 1 error is close to the nominal α while GRAMMAR is conservative and this conservativeness increases with the increase in the number of sibships and the heritability of the trait.

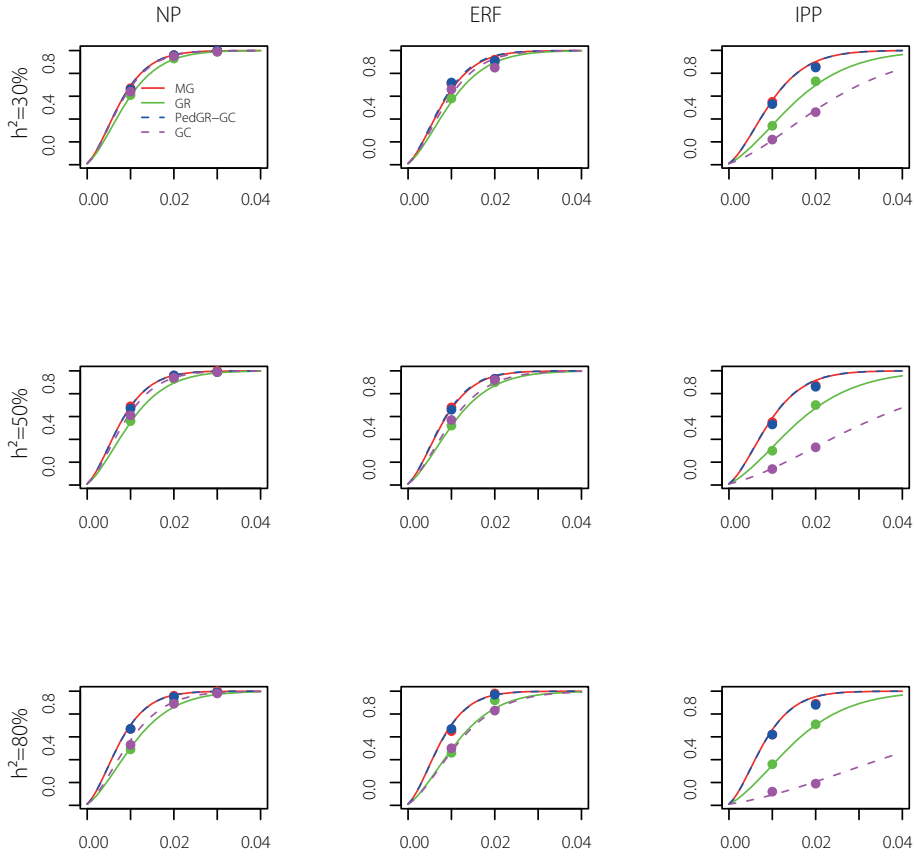


Figure 1: power of MG (red line), GRAMMAR (green line), PedGR-GC (blue dashed line), and GC (pink dashed line) to detect association under different heritability models and pedigree structures. The three rows show the power under different heritability models (from 30% to 80%) and the three columns show power achieved in different pedigrees namely nuclear pedigrees (NP), Erasmus Rucphen Family (ERF), and idealized pig population (IPP). The y-axis of each panel shows power while the x-axis shows the proportion of variance explained by the QTL under study. The red (for MG), green (for GRAMMAR), blue (for PedGR-GC), and pink (for GC) circles show the empirical power estimates. The power estimates are based on $\alpha = 0.01$. The empirical power estimates are based on 1000 simulations for NP, and IPP, and 100 simulations for ERF.

Supplementary Table 1 and **Figure 1** illustrate the power of the proposed (GC and PedGR-GC) and previous methods (MG and GRAMMAR). In the **Figure 1**, power to reach $\chi^2 \geq 6.64$ ($\alpha = 0.01$) is depicted with circles. From the available evaluation points we also estimated the slope of linear regression of non-centrality parameter on proportion of variance explained and used this slope to derive power curves.

Figure 1 shows that the power of PedGR-GC (blue dashed line) is very close to the power of the 'gold standard' measured genotype approach (red line) for all scenarios. These two methods appear to be the most powerful of all methods for all pedigree structures and genetic models evaluated. The power of GC (pink dashed line) is close to that of MG and PedGR-GC when the heritability is low but its' power declines when the heritability of the trait increases or when pedigrees with large number of full-sib families (IPP) are investigated. GRAMMAR (green line) performs similar to GC in a sample of nuclear families and in the ERF sample, but is more powerful when IPP pedigree is investigated.

To study the potential of described methods on genome-wide scale we used 695 people who are a part of the ERF pedigree and who were genotyped using Illumina 6K SNP linkage panel. Based on pedigree records, the 695 people formed 471 pairs of first-, 311 pairs of second-, 681 pairs of third- and 1,105 pairs of forth-degree relatives and 223,578 pairs of more distant relationship.

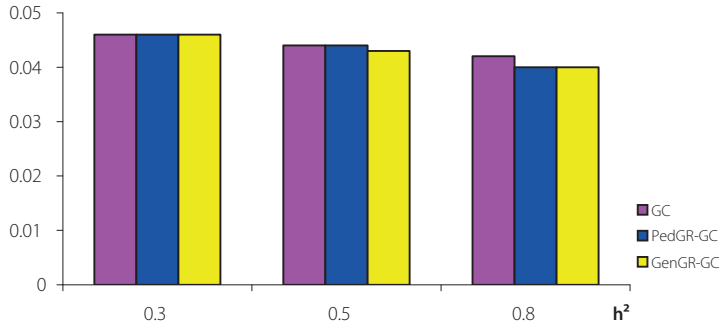
We generated 500 replicas for each of the models assuming total trait heritability of 30, 50 and 80%. In each replica, we selected five hundred random SNPs each explaining equal proportion of variance amounting to the total heritability minus 4%, and one random SNP explaining 4% of the phenotypic variance. The analysis trait was obtained as a sum of the SNP effects and a random number from the Normal distribution with mean zero and variance 0.7, 0.5 or 0.2 for total trait heritabilities of 30, 50 and 80%, respectively.

Type one error was estimated as the proportion of non-associated SNPs (> 2.5 million tests in total) showing association P-value of 5% or less. The statistical power was estimated as the proportion of replicas in which the SNP explaining 4% of variance passed genome-wide significant threshold (Bonferroni-corrected p-value $0.05 / 5524 = 9 * 10^{-6}$).

For analysis, we used GC, PedGR-GC and GenGR-GC. For GenGR-GC, the kinship matrix used was estimated from genomic information on 5524 autosomal SNPs typed in 695 members of the ERF study.

All methods showed a genome-wide type 1 error which was very close, but lower than the pre-specified threshold of 5% (**Figure 2A**). The methods tended to be more conservative at higher heritabilities. These results are consistent with the observations of others, that GC is conservative, and can be explained by the fact that all SNPs, some of which were associated with the trait, are used to estimate the null distribution of the test statistic.

(A) TYPE 1 ERROR



(B) POWER

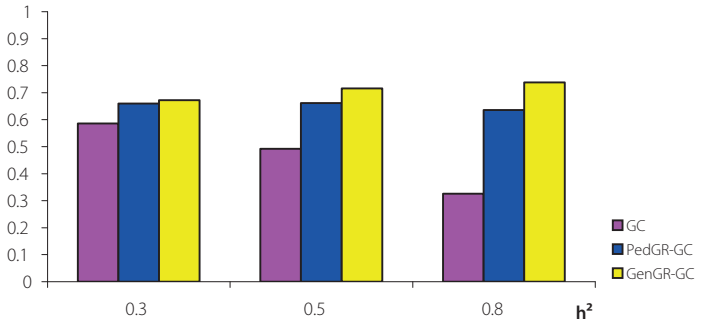


Figure 2: type 1 error (A) and power (B) to achieve 5% genome-wide significance at the truly associated SNP in a study of 695 ERF people genotyped for 5524 autosomal SNPs.

DISCUSSION

In this work we aimed to develop fast and powerful methods for genome-wide association analysis in samples of related individuals by exploiting the ideas of genomic background for correction of the distribution of the test statistics and for inferring the relation between study subjects. We show that methods, which exploit only genomic background, such as Genomic Control (GC) and GRAMMAR-GC using genomic kinship, are powerful and genome-wide feasible methods. Moreover, genomic GRAMMAR-GC, which infers genetic relations from genomic data, may be superior to the methods that use pedigree information in an exact manner.

Our simulations show that GC is a valid method to study data coming from samples of related individuals. GC can be a powerful tool for the analysis of pedigree based quantitative trait

loci. It outperforms traditional family based approaches like QTDT and FBAT (cf. Table 1 & Table 2 of Aulchenko et. al.[11]). The power of GC is close to that of the 'gold standard' measured genotype approach when trait heritability is low or moderate and human pedigrees are studied. However, the power of GC drops notably with high trait heritability and when pedigrees involve very large sibships, which are likely to be observed in animal pedigrees.

The results that GC was less powerful, than GRAMMAR-GC based methods, and tended to lose power at higher heritabilites (**Figure 2B**), are not surprising and are consistent with our previous findings [11]. Interestingly enough in a study of real Genome-wide data in ERF pedigree, GenGR-GC was consistently more powerful than pedigree-based GRAMMAR-GC (PedGR-GC), and the power advantage became more pronounced at higher heritabilities.

We proposed an extension to the GRAMMAR method [11], which increases its' power, maintains a nominal type 1 error and also does not require the precise knowledge of pedigree structure. Our method (GRAMMAR-GC) involves three steps which include removing the correlation from the data using relationship matrix estimated from either the pedigree or the genome (derivation of environmental residuals), using the correlation-free residuals from step 1 as the trait to perform association analysis, and then applying GC to correct the test statistic. We show through simulations that our proposed method performs very similar to the Measured Genotype (MG) approach with respect to type 1 error and power yet it is fast and feasible for genome-wide association analysis. By analysis of real genome-wide SNP data we showed that when the genomic data is used to estimate the relationship matrix (GenGR-GC) instead of the estimate obtained from genealogy, the power might be even improved.

One of our conclusions is that in genome-wide association studies of related individuals genomic background based methods such as genomic GRAMMAR-GC should be preferred over the ones exploiting known pedigree structure, such as pedigree GRAMMAR-GC or MG approach. There are two reasons why we believe that GenGR-GC should be preferred over its' pedigree analog.

First, errors in genealogy such as mis-specification of relations can lead to either false positives or negatives. Secondly, relationship coefficient computed from a pedigree is an expectation of the proportion of genome shared identical by descent (IBD) under the infinitesimal model, assuming infinite number of unlinked loci. The true proportion of genome shared, however, may deviate from this expectation [17]. For example, for remote relatives there is a fair chance of not sharing any genomic loci IBD. We may speculate that kinship estimated based on marker data can reflect the true unobserved genomic sharing better then the expectation computed from even a correct pedigree. If this is true, under the polygenic model we should expect that methods based on kinship estimated from marker data will be more powerful than the methods estimating kinship from the pedigree. We however leave a more detailed investigation of effects of pedigree error and precision of genomic kinship estimates on type 1 error and power to future works.

Another advantage of the GRAMMAR-GC is that the environmental residuals used for the analysis are free from familial correlations. Therefore the structure of the data becomes exchangeable. This means that permutation techniques may be applied to derive empirical measures of significance. In the analysis of data where adjacent SNPs are correlated due to linkage disequilibrium, thresholds set via permutation will account for these correlations and result in less stringent thresholds than those set by Bonferroni correction. This will become more and more important in the future, when denser marker maps with millions of SNPs will be applied to do association studies. Another attractive feature is that a range of new methods developed for classical "unrelated individuals" design can be applied to polygenic residuals obtained at the first stage of the analysis. In recent years, much progress was made in development of powerful methods and software which are robust to possible allelic heterogeneity through utilization of haplotype clustering and population genetic coalescence modeling [21-22].

Finally, GRAMMAR-GC is easily extendable: for example, it is easy to include covariates, interactions with sex and environment, gene-gene interactions and parent of origin effects.

To conclude, GRAMMAR-GC is a fast powerful approach for genome-wide association analysis of quantitative traits in samples of related individuals, which does not require precise knowledge of pedigree structure. This method is implemented as part of the GenABEL package, available at http://cran.r-project.org/.

***Acknowledgements*:**
We would like to thank Prof. Tatiana Axenovich, Prof. Haley, Dr de Koning, Dr. Catherine Bourgain, Prof. Ober, and Dr. William Astle for valuable discussions.

REFERENCES

1. Majumder, P.P., and Ghosh, S. (2005). Mapping quantitative trait loci in humans: achievements and limitations. J Clin Invest 115, 1419-1424.
2. Sing, C.F., and Davignon, J. (1985). Role of the apolipoprotein E polymorphism in determining normal plasma lipid and lipoprotein variation. Am J Hum Genet 37, 268-285.
3. Isaacs, A., Sayed-Tabatabaei, F.A., Aulchenko, Y.S., Zillikens, M.C., Sijbrands, E.J., Schut, A.F., Rutten, W.P., Pols, H.A., Witteman, J.C., Oostra, B.A., et al. (2007). Heritabilities, apolipoprotein E, and effects of inbreeding on plasma lipids in a genetically isolated population: the Erasmus Rucphen Family Study. Eur J Epidemiol 22, 99-105.
4. Risch, N., and Merikangas, K. (1996). The future of genetic studies of complex human diseases. Science 273, 1516-1517.
5. Abecasis, G.R., Cardon, L.R., and Cookson, W.O. (2000). A general test of association for quantitative traits in nuclear families. Am J Hum Genet 66, 279-292.
6. Horvath, S., Xu, X., Lake, S.L., Silverman, E.K., Weiss, S.T., and Laird, N.M. (2004). Family-based tests for associating haplotypes with general phenotype data: application to asthma genetics. Genet Epidemiol 26, 61-69.
7. Lange, C., DeMeo, D.L., and Laird, N.M. (2002). Power and design considerations for a general class of family-based association tests: quantitative traits. Am J Hum Genet 71, 1330-1341.
8. Boerwinkle, E., Chakraborty, R., and Sing, C.F. (1986). The use of measured genotype information in the analysis of quantitative phenotypes in man. I. Models and analytical methods. Ann Hum Genet 50, 181-194.
9. Lange, K., Sinsheimer, J.S., and Sobel, E. (2005). Association testing with Mendel. Genet Epidemiol 29, 36-50.
10. Havill, L.M., Dyer, T.D., Richardson, D.K., Mahaney, M.C., and Blangero, J. (2005). The quantitative trait linkage disequilibrium test: a more powerful alternative to the quantitative transmission disequilibrium test for use in the absence of population stratification. BMC Genet 6 Suppl 1, S91.
11. Aulchenko, Y.S., Koning, D.J., and Haley, C. (2007). GRAMMAR: a fast and simple method for genome-wide pedigree-based quantitative trait loci association analysis. Genetics.
12. Steinthorsdottir, V., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Jonsdottir, T., Walters, G.B., Styrkarsdottir, U., Gretarsdottir, S., Emilsson, V., Ghosh, S., et al. (2007). A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. Nat Genet 39, 770-775.
13. Devlin, B., and Roeder, K. (1999). Genomic control for association studies. Biometrics 55, 997-1004.
14. Bacanu, S.A., Devlin, B., and Roeder, K. (2002). Association studies for quantitative traits in structured populations. Genet Epidemiol 22, 78-93.
15. Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Serre, D., Boutin, P., Vincent, D., Belisle, A., Hadjadj, S., et al. (2007). A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445, 881-885.
16. Clayton, D.G., Walker, N.M., Smyth, D.J., Pask, R., Cooper, J.D., Maier, L.M., Smink, L.J., Lam, A.C., Ovington, N.R., Stevens, H.E., et al. (2005). Population structure, differential bias and genomic control in a large-scale, case-control association study. Nat Genet 37, 1243-1246.
17. Leutenegger, A.L., Prum, B., Genin, E., Verny, C., Lemainque, A., Clerget-Darpoux, F., and Thompson, E.A. (2003). Estimation of the inbreeding coefficient through use of genomic data. Am J Hum Genet 73, 516-523.
18. Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J., and Thompson, R. (2002). ASReml user guide release 1.0.(Orange, Australia: NSW Agriculture).
19. Aulchenko, Y.S., Ripke, S., Isaacs, A., and van Duijn, C.M. (2007). GenABEL: an R library for genome-wide association analysis. Bioinformatics.

20. Pardo, L.M., MacKay, I., Oostra, B., van Duijn, C.M., and Aulchenko, Y.S. (2005). The effect of genetic drift in a young genetically isolated population. Ann Hum Genet 69, 288-295.

21. Zollner, S., and Pritchard, J.K. (2005). Coalescent-based association mapping and fine mapping of complex trait loci. Genetics 169, 1071-1092.

22. Durrant, C., Zondervan, K.T., Cardon, L.R., Hunt, S., Deloukas, P., and Morris, A.P. (2004). Linkage disequilibrium mapping via cladistic analysis of single-nucleotide polymorphism haplotypes. Am J Hum Genet 75, 35-43.

SUPPLEMENTARY INFORMATION

Supplementary table 1. Empirical (permutation-based) type 1 error of disregarding-family-structure (DFS), 2-Step and FBAT analyses. Significantly inflated (conservative) type I errors are indicated in bold (*italic*), P≤0.05

Pedigree			α=0.05			α=0.01	
MAF	h^2	DFS	2-Step	FBAT	DFS	2-Step	FBAT
NP							
0.1	0.3	0.058	*0.029*	0.06	0.013	0.009	0.015
0.1	0.3q	0.06	*0.036*	0.061	**0.017**	0.005	0.01
0.1	0.5	**0.073**	*0.022*	0.048	0.015	*0.002*	0.01
0.5	0.3	**0.063**	0.039	0.052	**0.02**	0.005	0.012
0.5	0.3q	**0.062**	*0.032*	0.046	**0.018**	0.006	0.012
0.5	0.5	**0.085**	*0.028*	0.042	**0.019**	*0.003*	0.008
ERF							
0.1	0.3	**0.086**	*0.028*	0.049	0.02	0.005	0.005
0.1	0.3q	**0.077**	*0.031*	0.056	0.022	0.005	0.007
0.1	0.5	**0.108**	*0.023*	0.053	0.023	*0.003*	0.011

Supplementary Table 2. Mean percent of difference (95% distribution interval) and correlation (±s.e.m) between the values of χ^2 statistics coming from measured genotypes approach and 2-Step, disregarding-family-structure (DFS), QTDT and FBAT analyses at different pedigree structures (NF, ERF) and QTL explaining 1, 2 and 3% of total variance

		2-Step		DFS		QTDT		FBAT	
MAF	h^2	Difference	Corr.	Difference	Corr.	Difference	Corr.	Difference	Corr.
NP	qtl0.01								
0.1	0.3	-13.5 (-19.8;-7.4)	99.74±0.02	11.9 (-20.7;55)	98.08±0.12	-51.2 (-99.4;53.9)	63.54±1.85	-44.4 (-99.4;70.1)	63.51±1.85
0.1	0.3q	-13.6 (-19.9;-7.2)	99.74±0.02	9.7 (-24.1;44)	98.11±0.12	-51 (-98;56.9)	64.92±1.79	-42.8 (-97.7;83.3)	62.36±1.9
0.1	0.5	-20.5 (-27;-14.8)	99.74±0.02	13.8 (-39.9;80.1)	95.03±0.3	-53.6 (-96.9;23.7)	69.04±1.62	-40.7 (-95.8;55.6)	68.65±1.64
0.5	0.3	-13.6 (-19.4;-7.9)	99.8±0.01	11.5 (-21.3;51.8)	97.91±0.13	-54.7 (-97.9;39.6)	64.98±1.79	-48 (-97.2;63.1)	65.05±1.79
0.5	0.3q	-13.5 (-20.2;-7.8)	99.77±0.01	10.9 (-23.1;49.8)	97.83±0.13	-51.9 (-97.9;45.4)	59.35±2.01	-45.9 (-98.6;60)	58.55±2.04
0.5	0.5	-20.4 (-26;-15.2)	99.75±0.02	18.2 (-43.6;84.7)	94.72±0.32	-50.1 (-98.1;45.6)	68.61±1.64	-36.4 (-97.8;77.2)	68.96±1.63
NP	qtl0.02								
0.1	0.3	-13.7 (-20.6;-8)	99.55±0.03	9.9 (-11.8;35.2)	98.17±0.11	-58.7 (-94.2;-2.1)	63.63±1.85	-54.8 (-93.1;5.9)	62.67±1.89
0.1	0.3q	-13.9 (-19.9;-7.5)	99.63±0.02	9.8 (-10.3;36.4)	98.45±0.09	-59.6 (-93.8;-8.1)	64.67±1.81	-56 (-92.9;3.6)	60.84±1.96
0.1	0.5	-20.6 (-26.8;-15.1)	99.57±0.03	12.7 (-22.1;52.4)	95.47±0.27	-59.8 (-90.2;-18.3)	70.77±1.55	-50 (-87.1;0.7)	69.06±1.62
0.5	0.3	-14 (-20.3;-8)	99.62±0.02	10.8 (-9.6;36.8)	97.97±0.12	-61 (-93.6;-7)	66.78±1.72	-55.2 (-92.5;6.4)	66.69±1.72
0.5	0.3q	-13.9 (-20.2;-7.7)	99.55±0.03	11.1 (-8.7;34.5)	97.89±0.13	-61.8 (-93.9;-12.9)	66.38±1.73	-58 (-94.2;-3.1)	64.68±1.8
0.5	0.5	-20.5 (-26.1;-14.8)	99.58±0.03	12.5 (-22.6;50.7)	94.73±0.32	-59.3 (-90.8;-12.9)	69.89±1.58	-47.6 (-87.4;14)	69.56±1.6
NP	qtl0.03								
0.1	0.3	-14.3 (-21.4;-7.8)	99.31±0.04	10.2 (-4.5;28.5)	98.33±0.1	-61.7 (-89.1;-25)	66.01±1.75	-59 (-87.7;-21.2)	64.88±1.8
0.1	0.3q	-14.2 (-21.7;-7.6)	99.3±0.04	9.9 (-7.9;31.1)	98.11±0.11	-61 (-90.3;-18.3)	64.41±1.82	-58.6 (-90.1;-11.8)	59.88±1.99
0.1	0.5	-20.9 (-27.9;-15.2)	99.27±0.04	12.7 (-15.3;42.3)	95.36±0.28	-61 (-87.2;-27.4)	69.92±1.58	-53.1 (-83.5;-10.4)	67.55±1.69
0.5	0.3	-14.3 (-20.4;-8)	99.45±0.03	10 (-5.9;30.3)	98.21±0.11	-61.6 (-89.5;-19.9)	62.5±1.89	-56.6 (-87.4;-11.7)	62.4±1.9
0.5	0.3q	-14.3 (-21;-7.7)	99.25±0.05	10.6 (-6.2;29.6)	98.11±0.12	-62.3 (-90.6;-19.5)	62.65±1.89	-59.8 (-90.4;-15)	60.07±1.99
0.5	0.5	-20.9 (-26.4;-15.3)	99.41±0.04	13.2 (-15;45.1)	95±0.3	-61.5 (-88;-26.2)	70.83±1.54	-51.2 (-84.4;-5.6)	68.89±1.63
ERF	qtl0.01								
0.1	0.3	-15.4 (-23.7;-9.5)	99.74±0.05	15.5 (-40.5;85)	94.21±1.04	-66.2 (-99.7;22.5)	39.16±8.24	-65 (-98.7;44.1)	36.59±8.45
0.1	0.5	-21.7 (-29.9;-15.4)	99.64±0.07	23.1 (-60.1;169.9)	88.76±1.97	-61.7 (-99.9;19)	50.13±7.21	-59.3 (-99.9;12.4)	53.14±6.89
0.5	0.3	-16 (-21.3;-9.6)	99.76±0.04	14 (-44.3;95.6)	91.84±1.45	-58.7 (-99.6;93.5)	52.07±7.01	-55.7 (-99.7;106.9)	52.29±6.98
0.5	0.5	-22.7 (-29;-16.4)	99.48±0.1	31.7 (-74.1;138.5)	88±2.1	-51.8 (-99.9;146.8)	43.12±7.89	-34.8 (-99;141.4)	53.94±6.8
ERF	qtl0.02								
0.1	0.3	-15.7 (-22.2;-7.8)	99.52±0.09	12 (-30.5;58.7)	95.08±0.89	-66.8 (-97;-8)	57.99±6.34	-67.5 (-96.7;4.9)	52.73±6.93
0.1	0.5	-22.5 (-31.9;-15.4)	98.98±0.19	20.9 (-28.1;87.3)	82.35±3.01	-69.3 (-99.6;-19.3)	58.31±6.31	-68.1 (-99.2;-16.7)	54.67±6.72
0.5	0.3	-15.9 (-23;-9.9)	99.44±0.1	9.5 (-27;47.2)	93.79±1.11	-70.6 (-97.9;-14.7)	51.56±7.06	-69.8 (-98.5;-18.7)	52.99±6.91
0.5	0.5	-22 (-29;-16.2)	99.34±0.12	15.2 (-42.7;89.8)	85.1±2.57	-67.5 (-97.7;-13)	39.14±8.24	-64.6 (-98;-15.9)	42.82±7.92

Supplementary table 3. Power, based on empirical (permutation-based) threshold of significance for the 2-Step and FBAT analyses

		α=0.05		α=0.01	
MAF	h^2	2-Step	FBAT	2-Step	FBAT
NP	qtl0.01				
0.1	0.3	0.798	0.507	0.606	0.273
0.1	0.3q	0.832	0.509	0.603	0.285
0.1	0.5	0.794	0.572	0.569	0.336
0.5	0.3	0.83	0.533	0.645	0.285
0.5	0.3q	0.832	0.497	0.592	0.25
0.5	0.5	0.799	0.56	0.557	0.325
NP	qtl0.02				
0.1	0.3	0.991	0.803	0.934	0.603
0.1	0.3q	0.987	0.802	0.935	0.591
0.1	0.5	0.984	0.869	0.927	0.665
0.5	0.3	0.988	0.805	0.942	0.619
0.5	0.3q	0.988	0.782	0.948	0.539
0.5	0.5	0.973	0.868	0.918	0.668
NP	qtl0.03				
0.1	0.3	0.999	0.934	0.992	0.806
0.1	0.3q	0.997	0.923	0.991	0.793
0.1	0.5	0.999	0.956	0.986	0.863
0.5	0.3	1.000	0.945	0.999	0.812
0.5	0.3q	1.000	0.916	0.992	0.781
0.5	0.5	0.999	0.959	0.989	0.857
ERF	qtl0.01				
0.1	0.3	0.78	0.32	0.6	0.14
0.1	0.5	0.76	0.39	0.54	0.19
0.5	0.3	0.84	0.34	0.52	0.12
0.5	0.5	0.69	0.29	0.47	0.11
ERF	qtl0.02				
0.1	0.3	0.96	0.54	0.89	0.3
0.1	0.5	0.98	0.61	0.90	0.41
0.5	0.3	0.95	0.51	0.90	0.24
0.5	0.5	0.99	0.69	0.90	0.42

Chapter 2

Attention Deficit Hyperactivity Disorder (ADHD)

Chapter 2.1

Suggestive linkage of ADHD to chromosome 18q22 in a young genetically isolated Dutch population

N. Amin[1], Y.S. Aulchenko[1], M.C. Dekker[2], R.F. Ferdinand[3], A. van Spreeken[4], A. H. Temmink[5], F.C. Verhulst[6], B.A. Oostra[1], C.M. van Duijn[1*]

[1] Genetic Epidemiology Unit, Department of Epidemiology & Biostatistics and Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands
[2] Department of Neurology, University Medical Centre Nijmegen, The Netherlands
[3] De Waag Center for Forensic Psychiatry, Rotterdam, The Netherlands
[4] Department of Neurology, Sint Franciscus Hospital, Roosendaal, The Netherlands
[5] Department of Neurology, Amphia Hospital, Breda, The Netherlands
[6] Department of Child and Adolescent Psychiatry, Erasmus MC Sophia, Rotterdam, The Netherlands

ABSTRACT

Attention Deficit/Hyperactivity Disorder (ADHD) is a common, highly heritable, neuropsychiatric disorder among children. Linkage studies in isolated populations have proved powerful to detect variants for complex diseases, such as ADHD.

We performed a genome-wide linkage scan for ADHD in 9 patients from a genetically isolated population in the Netherlands, who were linked to each other within ten generations through multiple lines of descent. The genome-wide scan was performed with a set of 400 microsatellite markers with an average spacing of ± 10-12 centimorgan. We performed multipoint parametric linkage analyses using both recessive and dominant models. Our genome scan pointed to several chromosomal regions that may harbour ADHD susceptibility genes. None exceeded the empirical genome-wide significance threshold but the LOD scores were >1.5 for regions 6p22 (HLOD = 1.67), and 18q21-22 (HLOD = 2.13) under a recessive model. We followed up these two regions in a larger sample of ADHD patients (n = 21, 9 initial and 12 extra patients). The LOD scores did not increase after increasing the sample size [6p22 (HLOD = .51, 18q21-22 (HLOD = 1.83)]. However, the LOD score on 6p22 increased to 2 when a separate analysis was performed for the inattentive type ADHD children. The linkage region on chromosome 18q overlaps with the findings of association of rs2311120 (p-value = 10^{-5}) and rs4149601 (p-value = 10^{-4}) in the genome-wide association analysis for ADHD performed by the GAIN consortium. Further, there was an excess of regions harbouring serotonin receptors (*HTR1B*, *HTR1E*, *HTR4*, *HTR1D*, and *HTR6*) that showed a LOD score > 1 in our genome-wide scan.

INTRODUCTION

Attention deficit/hyperactivity disorder (ADHD) is a disruptive behaviour disorder characterized by a persistent pattern of inattention and/or hyperactivity-impulsivity that is more frequent and severe than is typical for unaffected individuals in the same stage of development. Some impairment from these symptoms must be present in at least two settings, e.g. at home and at school[1]. ADHD has an onset in childhood but it can persist through adolescence and into adulthood. Five to seven percent of children and three percent of adults are estimated to be affected with ADHD[2-3]. Boys are three to four times more often diagnosed with ADHD than girls[3-4].

ADHD is a complex disorder influenced by both genetic and environmental factors. Heritability estimates from twin and adoption studies show a strong genetic component ranging from 60 to 90%[5-10], and sibling relative risk estimates range from a 4 to an 8 fold increase[11-12].

A large number of genetic studies of ADHD followed a candidate gene approach focusing mainly on the genes involved in the dopaminergic and serotonergic pathways[13]. Genes studied most are the dopamine transporter gene (*DAT1*), which maps to 5p15, the dopamine D4 (*DRD4*, 11p15) and D5 (*DRD5*, 4p16) receptor genes. These studies have yielded a number of replicated findings but meta- analyses show that the associated variants are of small effect sizes, with odds ratio ranging from 1.13 to1.9[14-15].

Until now, there have been 7 independent genome-wide linkage scans for ADHD. These include affected sib-pair (ASP) linkage studies[16-24] and studies of extended multigenerational families[25-26]. These studies suggested linkage to 1p36, 2q21, 2q35, 4q13.2, 5p13, 5q13.1, 5q33.3, 6q12, 6q22-23, 7p13, 7q21, 9q22, 11q22, 13q12, 14q12, 15q15, 16q23, 17p11, and several other regions with nominally significant evidence of linkage but no outstanding replications. The continued failure to replicate linkage findings for ADHD has led researchers to believe that the genes affecting ADHD have common variants with very small effects that cannot be detected successfully with methods relying on linkage and hence advocated the use of association analysis[20], or alternatively, are rare variants with strong effects that fail to replicate because of family specific mutations. A recent genome-wide association study of ADHD performed by the International Multisite ADHD Genetics (IMAGE) group, which was conducted as a part of Genetic Association Information Network (GAIN), included 958 parent-child trios and 600,000 SNPs, but failed to provide convincing evidence for a number of common risk variants[27].

While association analysis is a powerful tool to detect common variants with small effects, linkage analysis has proven successful in the detection of rare variants with large effects. Linkage, for common diseases, has been very successful in isolated populations[28-29], as drift and founder effects lead to the extinction of most rare variants, while a small number is retained, which, over subsequent generations, become frequent[30]. This is crucial for linkage as it implies that genetic heterogeneity is reduced[31-32].

In the current study, we report the results of an independent genome-wide linkage scan of ADHD children, which was performed in a genetically isolated population in the Netherlands. We compare our linkage findings to that of the genome-wide association study of ADHD performed by the IMAGE group.

MATERIAL & METHODS

Study Population

This study was conducted within the framework of the *Genetic Research in Isolated populations (GRIP)* program. Approximately 150 individuals founded this population, located in the South West of The Netherlands, in the middle of the 18th century. The population expanded from 700 inhabitants in 1848 to more than 20,000 inhabitants at present. For this population, a genealogical database including records for more than 100,000 individuals is available.

For this study, two paediatric neurologists, to whom ADHD patients are referred in GRIP, asked all of their patients diagnosed with ADHD to participate in this study (n = 49, 22% females)[33]. Thirty-three (67%) patients and their parents agreed to participate. Of these 33 patients 2 were excluded from analysis because their genealogy could not be worked out, and 5 children were excluded because they did not fulfil the criteria used for the diagnosis of ADHD in the present study. Of the remaining 26 patients, 21 were inbred, of whom only 9 patients, who could be linked to each other within no more than 10 generations, were used in the initial analysis, and all 21 inbred patients were used in the follow up analysis.

Table1: Baseline characteristics of the study population

Characteristics	All	Included in genome-wide linkage analysis
Number of patients	26	9
Number of inbred patients	21	9
Number of patients who received diagnosis from two informants	16	7
Mean age at examination (range)	10(6-16)	10(6-15)
Females (%)	23.1	33.3
Mean Kinship	0.002	0.014
Mean Inbreeding	0.002	0.006
ADHD subtype		
Inattentive	12	5
Hyperactive-impulsive	3	-
Combined	11	4

Psychiatric Assessment

The Dutch version of the National Institute of Mental Health Diagnostic Interview Schedule for Children (NIMH DISC or DISC)-IV was used to assess DSM-IV diagnoses[34-36]. Psychologists and psychology students trained by the authors of the Dutch DISC-IV administered the DISCs. The training schedule used was similar to that used by the authors of the original English version, at Columbia University, New York. To obtain information regarding a wide range of current DSM-IV Axis 1 diagnoses, parent DISCs (DISC-P) were administered during face-to-face contacts, at a community general health centre or in a children's hospital. Furthermore, lifetime ADHD symptoms were also assessed with the DISC-P. Teachers were interviewed with the ADHD section (current, not lifetime) of the teacher DISC (DISC-T) via telephone. The child version of the DISC (DISC-C) was not applied since most of the children included in our sample were too young (< 11 years of age). Children receiving treatment were withdrawn from medication for this study prior to the interview.

Phenotypic subgroups (inattentive, hyperactive/impulsive, and combined) of ADHD were formed based on application of the DSM-IV criteria that had been assessed with the DISC. Current ADHD diagnoses were based on information from both parents and teachers. Two types of ADHD diagnoses were derived: (1) 'based on one informant', and (2) 'based on two informants'. A diagnosis of ADHD based on one informant was applied when either a parent or a teacher scored six or more criteria for the inattentive, hyperactive or combined phenotype positive, while the other informant scored less than three criteria positive. A diagnosis of ADHD based on two informants was applied when one informant scored six or more criteria of one of the ADHD subgroups positive and the second informant scored three or more criteria positive. The threshold of '3 criteria positive' was chosen arbitrarily for the purpose of the present study. The DSM-IV does not provide explicit rules for the number of criteria that need to be positive in 2 settings to obtain an ADHD diagnoses. It merely states that symptoms have to be present in at least 2 settings. If a child did not fulfill the criteria for current ADHD with the DISC-P, lifetime information from the DISC-P was used to obtain a lifetime diagnosis of ADHD. The baseline characteristics of the patients are given in **table 1.**

Genotyping

Blood was drawn for all patients and their parents. DNA was extracted from peripheral leucocytes using standard procedures[37]. We performed the genome-wide linkage scan on 9 patients and their parents with a set of 400 fluorescently labelled, highly polymorphic micro-satellite markers (distance between markers ± 10-12 cM) covering the whole genome. The remaining, distantly related, 17 patients and their parents were only typed for the markers in the regions of interest on chromosomes 6 (n = 12) and 18 (n = 7). The genotyping experiments were done following manufacturer instructions (Applied Biosystems, Foster City, CA).

Statistical Analysis

For the 9 patients used in the initial genome-wide scan, a pedigree was extracted from the database. Since the size of the pedigree was too large (n = 2206) to be analyzed with any of the available soft ware packages for linkage analysis, we divided the pedigree into two smaller, analysable sub-pedigrees using PEDCUT software. This program allows identification of sub-pedigrees that fall within a pre-specified pedigree bit-size limit that can be analyzed, maximizing the size of the subgroups[38]. One of sub-pedigrees (**Figure 1a**) had bit-size of 46 and contained 146 individuals (6 patients) and the other sub-pedigree (**Figure 1b**) had bit-size of 19 and contained 41 individuals (3 patients).

We checked all the markers for Mendelian inconsistencies using PEDCHECK, and in the case of inconsistencies, a second round of laboratory quality control was performed. In case the reasons for the problem could not be identified, the genotypes of the parents and child(ren) were set to missing. Marker allele frequencies were estimated by pooling the data from a sample of 447 people from the same population, using the maximum likelihood method as implemented in the PoolSTR software[39]. Data handling and preparation of input files was done with MEGA2[40].

For the genome-wide linkage analysis, we performed affected only analyses using both dominant and the recessive models. Multipoint parametric linkage analysis under the dominant model was performed assuming a disease allele frequency of 0.001, complete penetrance and a phenocopy rate of 0.01 using SIMWALK2.

The recessive analysis was performed using homozygosity mapping[41]. We adjusted for inbreeding using the shortest loop and a hypothetical loop capturing all cryptic inbreeding[42]. The disease allele frequency was set to 0.01. A model with complete penetrance and no phenocopies was used to perform heterogeneity LOD score computations with MERLIN[43]. Haplotypes were constructed using SIMWALK2.

The genome-wide significance thresholds were determined empirically by performing 1000 genome-wide simulations of our data under the null hypothesis of no linkage. We used the complete pedigree, including all 2206 members, for marker simulation. Unlinked markers were dropped in the complete pedigree. Number of markers and intermarker distances were simulated according to the typed marker set. We performed linkage analysis using the subpedigrees. Disease-allele frequency, genetic models, pedigrees, and penetrances were the same as those we used in the actual linkage analysis. Genotypes of untyped individuals were set to "missing." For each genome screen, the highest HLOD score was recorded. The cumulative density function of the obtained 1000 maximum HLOD scores approximates the distribution of the genome-wide type I error rates. Our simulations showed that an HLOD score of 2.65 corresponds to a genome-wide type I error rate of 5% and that an HLOD of 1.78 corresponds to a genome-wide type I error of 50%.

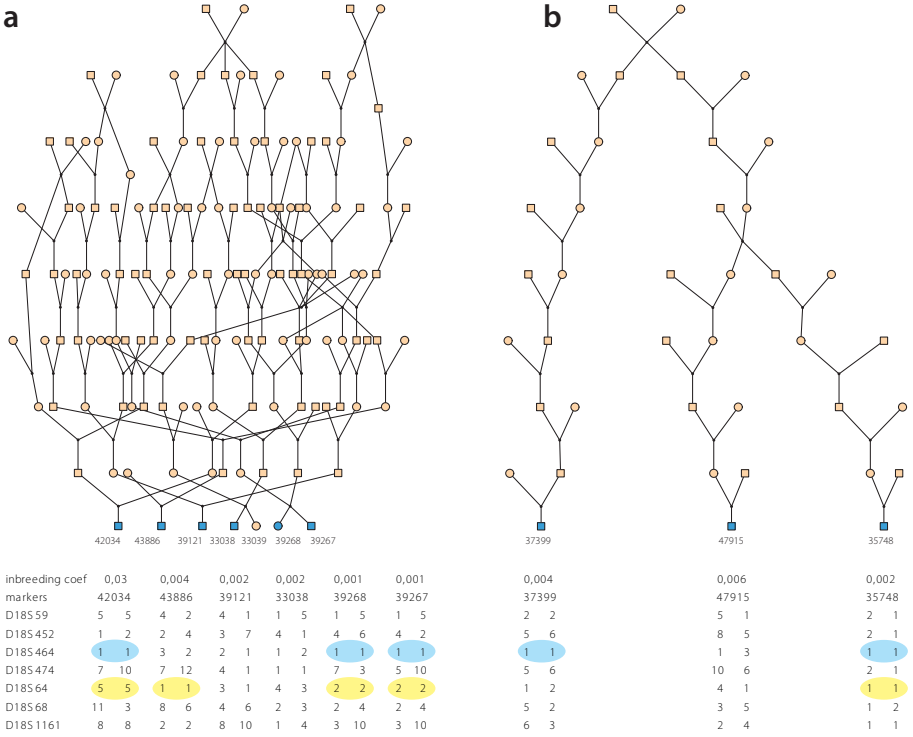


Figure 1: sub-pedigrees after breaking the complete pedigree. Inbreeding coefficients and haplotypes for chromosome 18 for all 9 patients are shown at the bottom

RESULTS

As expected, due to the selection on relationship within 10 generations, those included in the initial genome-wide linkage analysis had an almost 10 fold higher average kinship compared to that of all patients. These patients also showed an increased inbreeding coefficient, which may point to a recessive form of disease.

Results of the complete genome wide scan from both the dominant model and recessive model with shortest and hypothetical loops are illustrated in the **Figures 2a**, **2b**, and **2c**.

There was no genome-wide significant evidence for linkage under either model. The highest LOD score under the dominant model was observed at 6q16 (HLOD = 0.91). Other regions that showed weak evidence of linkage include 2q23-24 (HLOD = 0.81), 3q24 (HLOD = 0.75), and 12p13 (HLOD = 0.71). Homozygosity mapping yielded 5 genomic regions with HLOD ≥ 1. The strongest evidence of linkage was observed at 18q21-22 (D18S64, HLOD = 2.13). Other regions with HLOD ≥ 1 include 6p23 (D6S470, HLOD = 1.68), 6p12 (D6S257, HLOD = 1.07),

1p36 (D1S214, HLOD = 1.09), 18p11 (D18S59, HLOD = 1.15), and 15q25 (D15S205, HLOD = 1.19). Details are provided in table 2. Adjusting for multiple inbreeding loops (**Figure 2c**) did not alter our findings, decreasing LOD scores only marginally.

Table 2: MLS ≥ 1 under homozygosity mapping

Chromosome	Position (cM)	Most likely cytogenetic location	Nearest marker	Multipoint HLOD based on homozygosity mapping	
				Liberal*[1]	Conservative*[2]
1	17	1p36	D1S450	1.1	0.258
5	174	5q33	D5S422	0.73	1.03
6	20	6p22	D6S289	1.67	0.85
6	69	6p12	D6S257	1.07	0.47
6	102	6q15	D6S462	1.125	0.58
15	78.5	15q25	D15S205	1.19	0.54
18	117	18q21	D18S64	2.13	1.27
18	140	18q22	D18S1161	1.2	1.81
18	0	18p11	D18S59	1.15	0.22

*estimated using MERLIN

[1] results from the linkage analysis of pedigrees based on shortest inbreeding loop.

[2] results from the linkage analysis of pedigrees constructed hypothetically using the inbreeding coefficient

The patients' haplotypes at chromosome 18 (presented in **Figure 1**) show excess of homozygosity but not at a single marker. Four out of nine patients are homozygous for allele 1 of the marker D18S464. This is, however, the most common allele, with a frequency of homozygosity of 0.42. Also, at marker D18S64, five patients are homozygous: two are homozygous for allele 1 which has a frequency of homozygosity of 0.05, two are homozygous for allele 2 which has a frequency of homozygosity of about 0.002 and one patient is homozygous for allele 5 which has a frequency of homozygosity of 10^{-9}.

The regions of interest on chromosome 18 and 6 were additionally typed for the remaining distantly related 17 patients, and the data from all 21 inbred patients (**Figure3**) was reanalyzed. For this analysis, we used only the recessive model of inheritance as it yielded the evidence for linkage in the initial analyses.

The LOD scores did not increase by increasing the sample size, but rather decreased due to adding non-informative individuals (**Figure 4**). On chromosome 6p, the highest LOD score we observed was HLOD = 1.51, at marker D6S470, using homozygosity mapping. This HLOD score increased to 2 at marker D6S1574 when a recessive model with a disease allele frequency of 0.01 and complete penetrance was run separately for the 10 patients with only inattentive type ADHD (**Figure 4**). For chromosome 18, the highest LOD score observed was

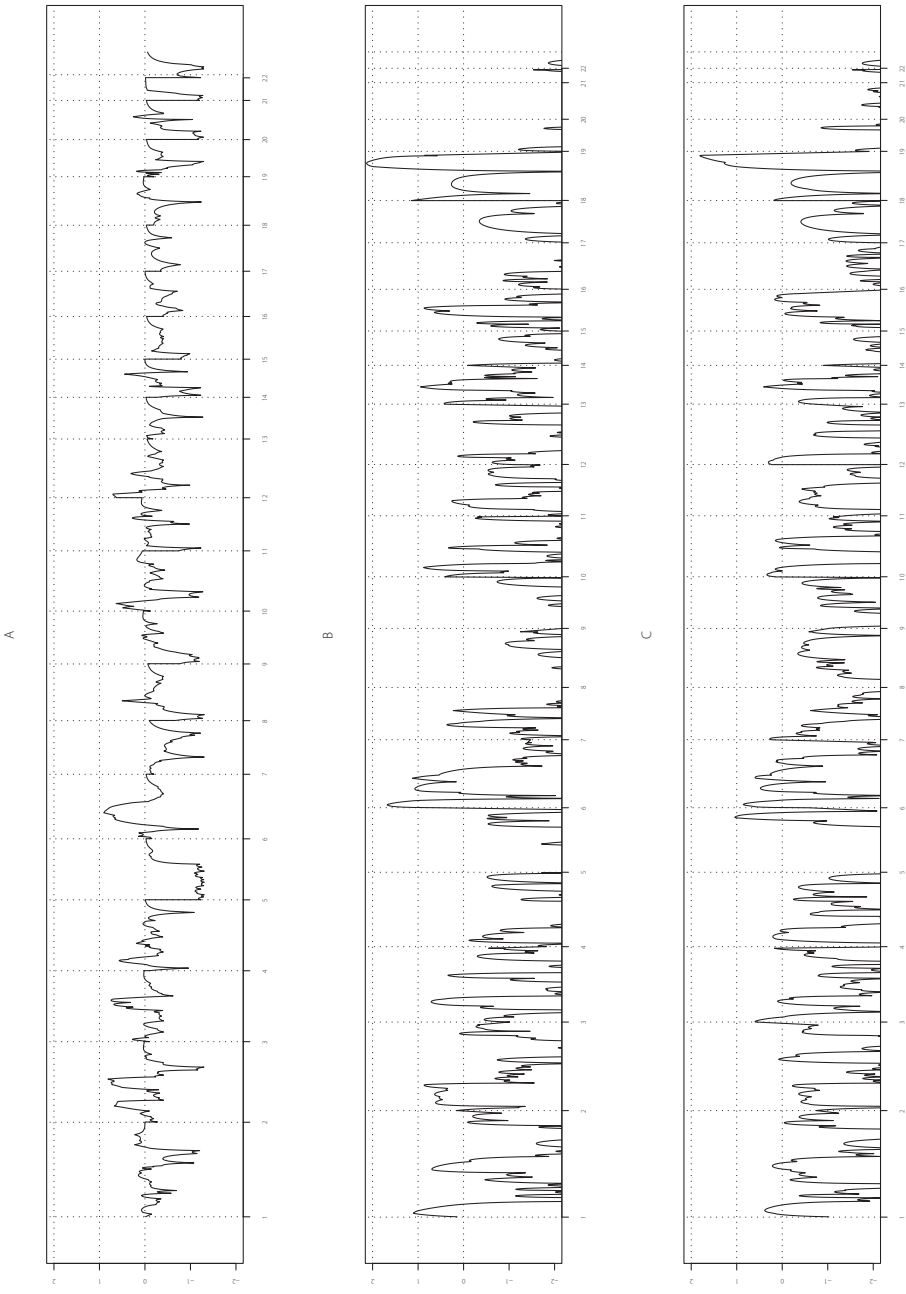


Figure 2: LOD score plots from multipoint analyses of the whole autosomal genome in 9 ADHD patients. A shows the LOD scores for dominant model, B and C for recessive model under homozygosity mapping with the shortest loop and hypothetical loop respectively. The horizontal axis depicts the whole genome divided into 22 autosomes, and the vertical axis depicts the LOD scores in each panel.

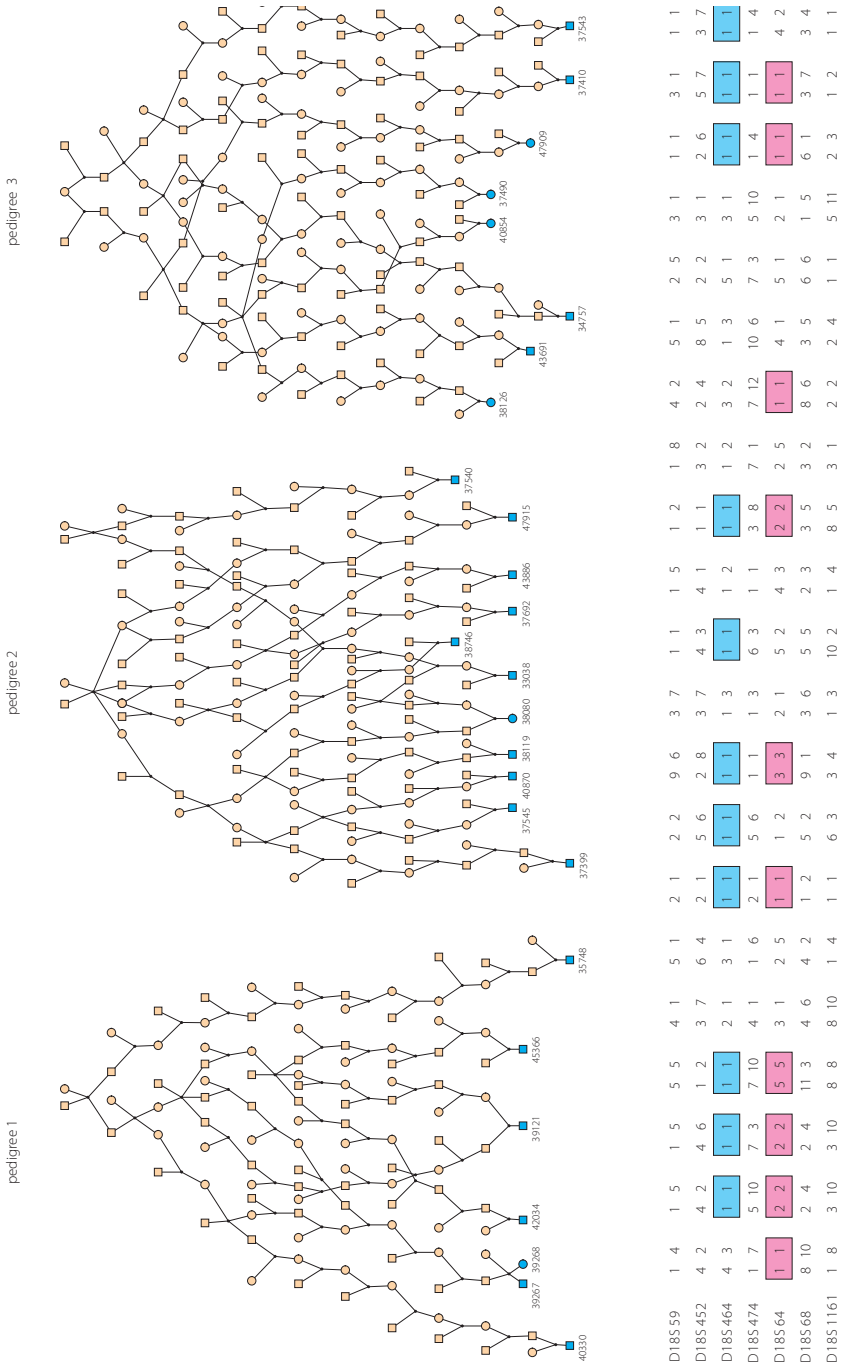


Figure 3: sub-pedigrees after breaking the complete pedigree for all 26 patients and haplotypes for chromosome 18 for only the inbred patients (n = 21) are shown at the bottom.

1.83 at marker D18S1161. The haplotype analysis, however, revealed that 11 out of 21 inbred patients were homozygous for the same allele at marker D18S464 (**Figure 3**) and 10 patients were homozygous at marker D18S64.

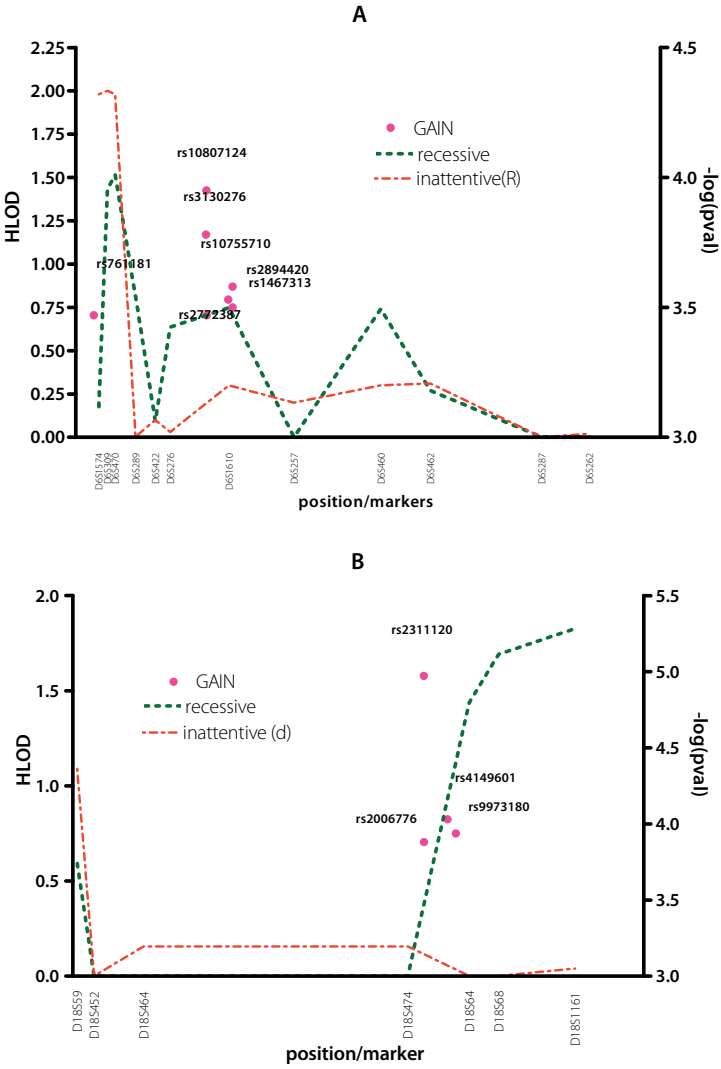


Figure 4: HLOD score plots from the follow-up multipoint analyses with increased sample size for chromosomes 6 and 18 and also the top GAIN hits for the same regions. The horizontal axis depicts the distance in base pairs, the left vertical axis depicts the LOD/HLOD scores from the linkage analysis and the right vertical axis shows the –log10(p-value) for the association analysis. The pink dots show the top hits of GAIN study in the region, the green line depicts the results from recessive model using homozygosity mapping, and the red dotted line in 4A shows the results of a recessive model run using SIMWALK2 for the inattentive subtype only and in 4B shows the results of a dominant model that was run separately for the inattentive type only using SIMWALK2.

Finally, we compared our findings to those from the genome-wide association analysis of the GAIN consortium. The chromosome 18 region identified in our linkage analysis also showed evidence for association in the GAIN analysis; rs2311120 (p-value = 10^{-5}), rs4149601 (p-value = 10^{-4}), rs9973180 (p-value = $10^{-4)}$ and rs2006776, (p-value = 10^{-4}) are located in the linkage region we identified (**Figure 4**). These single nucleotide polymorphisms (SNPs), particularly rs2311120, which was the third most significant SNP, were among the top 100 most strongly associated SNPs in the GAIN consortium. Also for chromosome 6p in the region we identified in our genome-wide scan there was also some evidence of association in the GAIN study (rs2772387, p-value = $3*10^{-4}$).

DISCUSSION

We performed a genome-wide scan using two extended families from a genetically isolated Dutch population. Since these pedigrees were selected based on consanguinity, we selected families with recessive form of disease. There was an increased inbreeding so we did not expect a dominant form of disease. We, however, tested a dominant model because in isolated populations the disease may appear in pseudo-dominant forms. We did not observe evidence of significant or even suggestive linkage under a dominant model. Using a recessive model, we identified 6 genomic regions with HLOD score >1. Although none of these regions passed the genome-wide significance threshold, there was suggestive evidence of linkage at 18q21-22 (HLOD = 2.13, marker D18S64). The LOD score did not increase with the increase in the sample size. This may be explained by the fact that the patients selected for the genome-wide linkage analysis were the ones showing the strongest evidence for a recessive form of the disease, as the inbreeding coefficient was the highest for these patients (**Table 1**). The region 18q21-22 has been implicated earlier as a major susceptibility locus for bipolar disorder[44]. None of the children in our sample received a clinical or DISC-P diagnosis of bipolar disorder, and none of the children received a clinical diagnosis of schizophrenia or were scored positive on the psychosis screen of the DISC-P. Since none of the patients included in our study showed evidence of bipolar disorder or schizophrenia, there is no evidence of misdiagnosis explaining our finding. Considering the fact that some of the symptoms of ADHD and bipolar disorder coincide, and that ADHD in childhood increases the risk for later developing bipolar disorder, this finding suggests that this locus might harbour pleiotropic genes that increase the risk of both ADHD and bipolar disorder. Our findings are further supported by the fact that this region showed evidence of association in the GAIN study; rs2311120 was the third most significant SNP in the GAIN analysis. There were also other, less significant, SNPs in the same region (rs9973180, p-value = 10^{-4}, rs2006776, p-value = 10^{-4}) supporting the hypothesis that this region is implicated in ADHD. This region, however, did not show evidence of linkage when an ASP linkage scan[19] was performed by the IMAGE group, which used a sample largely from the GAIN families.

The other interesting region in our genome-wide scan comprises of two adjacent regions on chromosome 6: 6p22-24 (HLOD = 1.67, marker D6S257) and 6p12 (HLOD = 1.07, marker D6S257). 6p22 is one of the most frequently replicated susceptibility regions for reading disability (RD) or dyslexia[45-48], and has also been implicated as a susceptibility locus for ADHD[49] in a study of ADHD within sibpairs identified for RD . In a sub-analysis of this region, the LOD score increased to HLOD = 2 (marker D6S309) when the data were analysed separately for the patients having predominantly inattentive type ADHD. However, subgroup analyses may cause false positive findings and remain to be confirmed. When considering other regions that were suggested by others, the region 6q15 that had a HLOD of 1.13 in our study and also harbours serotonin receptor genes *HTR1B* and *HTR1E* is adjacent (distance = 10 cM) to the region 6q14, which was identified by Ogdie et al[23-24] as a nominally significant susceptibility locus for ADHD. Our genome scan also showed some evidence of linkage to 5q33 (conservative homozygosity mapping HLOD = 1.03, marker D5S422) (**Figure 2c**). This region harbours serotonin receptor 4 (*HTR4*) and was first identified with significant evidence of linkage in a genome-wide scan for ADHD in an isolated population from Colombia[25]. The region 1p36 (HLOD = 1.09, marker = D1S450) has recently been identified as a susceptibility locus for ADHD with significant evidence of linkage in a linkage study of quantitative ADHD traits performed by the IMAGE group[22]. This region harbours serotonin receptor 6 (*HTR6*), and serotonin receptor 1D (*HTR1D*) genes. It is of interest that each of these marginally linked regions includes serotonin receptors. *HTR1B*, *HTR1E*, *HTR4*, *HTR1D*, and *HTR6* genes are known candidate genes for ADHD, although, except for *HTR1B*[50-52], the association of these genes with ADHD has not been established.

Although our study sample was small, the strength of our population lies in the fact that we can select patients based on genealogy. In that way, we can specifically target patients with dominant, or, in this case, recessive forms of disease based on their consanguinity. Taken together with the GAIN results, our study yields evidence that 18q21-22 may be relevant for ADHD. Our findings, in conjunction with those of GAIN, ask for further follow-up of the region. Further, our study suggests that the serotonin receptors *HTR1B*, *HTR1E*, *HTR4*, *HTR1D*, and *HTR6* might be implicated in ADHD.

While isolated populations may facilitate the detection of linkage, caution is required in generalising the results to other populations. It is, therefore, necessary that these regions be followed up in other populations. Although our findings are compatible with those of GAIN, their credibility will increase if confirmed elsewhere.

Acknowledgement

We acknowledge the collaboration of all patients, their parents, and their teachers. We further thank Petra Veraart for the genealogy. We also thank Aaron Isaacs for verifying the manuscript. This study was supported by a grant from the Sophia Foundation for Scientific Research (SSWO, # 318) and the Center of Medical Systems Biology (CMSB).

REFERENCES

1. (1994). American Psychiatric Association.(Washington DC).
2. Wolraich, M.L., Hannah, J.N., Pinnock, T.Y., Baumgaertel, A., and Brown, J. (1996). Comparison of diagnostic criteria for attention-deficit hyperactivity disorder in a county-wide sample. J Am Acad Child Adolesc Psychiatry 35, 319-324.
3. Swanson, J.M., Sergeant, J.A., Taylor, E., Sonuga-Barke, E.J., Jensen, P.S., and Cantwell, D.P. (1998). Attention-deficit hyperactivity disorder and hyperkinetic disorder. Lancet 351, 429-433.
4. Cantwell, D.P. (1996). Attention deficit disorder: a review of the past 10 years. In J Am Acad Child Adolesc Psychiatry. pp 978-987.
5. Thapar, A., Hervas, A., and McGuffin, P. (1995). Childhood hyperactivity scores are highly heritable and show sibling competition effects: twin study evidence. Behav Genet 25, 537-544.
6. Edelbrock, C., Rende, R., Plomin, R., and Thompson, L.A. (1995). A twin study of competence and problem behavior in childhood and early adolescence. J Child Psychol Psychiatry 36, 775-785.
7. Faraone, S.V., and Doyle, A.E. (2001). The nature and heritability of attention-deficit/hyperactivity disorder. Child Adolesc Psychiatr Clin N Am 10, 299-316, viii-ix.
8. Gjone, H., Stevenson, J., and Sundet, J.M. (1996). Genetic influence on parent-reported attention-related problems in a Norwegian general population twin sample. J Am Acad Child Adolesc Psychiatry 35, 588-596; discussion 596-588.
9. Levy, F., Hay, D.A., McStephen, M., Wood, C., and Waldman, I. (1997). Attention-deficit hyperactivity disorder: a category or a continuum? Genetic analysis of a large-scale twin study. J Am Acad Child Adolesc Psychiatry 36, 737-744.
10. Sherman, D.K., Iacono, W.G., and McGue, M.K. (1997). Attention-deficit hyperactivity disorder dimensions: a twin study of inattention and impulsivity-hyperactivity. J Am Acad Child Adolesc Psychiatry 36, 745-753.
11. Smalley, S.L. (1997). Genetic influences in childhood-onset psychiatric disorders: autism and attention-deficit/hyperactivity disorder. Am J Hum Genet 60, 1276-1282.
12. Faraone, S.V., Biederman, J., and Monuteaux, M.C. (2000). Toward guidelines for pedigree selection in genetic studies of attention deficit hyperactivity disorder. Genet Epidemiol 18, 1-16.
13. Thapar, A., O'Donovan, M., and Owen, M.J. (2005). The genetics of attention deficit hyperactivity disorder. Hum Mol Genet 14 Spec No. 2, R275-282.
14. Faraone, S.V., Doyle, A.E., Mick, E., and Biederman, J. (2001). Meta-analysis of the association between the 7-repeat allele of the dopamine D(4) receptor gene and attention deficit hyperactivity disorder. Am J Psychiatry 158, 1052-1057.
15. Maher, B.S., Marazita, M.L., Ferrell, R.E., and Vanyukov, M.M. (2002). Dopamine system genes and attention deficit hyperactivity disorder: a meta-analysis. Psychiatr Genet 12, 207-215.
16. Fisher, S.E., Francks, C., McCracken, J.T., McGough, J.J., Marlow, A.J., MacPhie, I.L., Newbury, D.F., Crawford, L.R., Palmer, C.G., Woodward, J.A., et al. (2002). A genomewide scan for loci involved in attention-deficit/hyperactivity disorder. Am J Hum Genet 70, 1183-1196.
17. Bakker, S.C., van der Meulen, E.M., Buitelaar, J.K., Sandkuijl, L.A., Pauls, D.L., Monsuur, A.J., van 't Slot, R., Minderaa, R.B., Gunning, W.B., Pearson, P.L., et al. (2003). A whole-genome scan in 164 Dutch sib pairs with attention-deficit/hyperactivity disorder: suggestive evidence for linkage on chromosomes 7p and 15q. Am J Hum Genet 72, 1251-1260.
18. Hebebrand, J., Dempfle, A., Saar, K., Thiele, H., Herpertz-Dahlmann, B., Linder, M., Kiefl, H., Remschmidt, H., Hemminger, U., Warnke, A., et al. (2006). A genome-wide scan for attention-deficit/hyperactivity disorder in 155 German sib-pairs. Mol Psychiatry 11, 196-205.
19. Asherson, P., Zhou, K., Anney, R.J., Franke, B., Buitelaar, J., Ebstein, R., Gill, M., Altink, M., Arnold, R., Boer, F., et al. (2008). A high-density SNP linkage scan with 142 combined subtype ADHD sib pairs identifies linkage regions on chromosomes 9 and 16. Mol Psychiatry 13, 514-521.

20. Faraone, S.V., Doyle, A.E., Lasky-Su, J., Sklar, P.B., D'Angelo, E., Gonzalez-Heydrich, J., Kratochvil, C., Mick, E., Klein, K., Rezac, A.J., et al. (2007). Linkage analysis of attention deficit hyperactivity disorder. Am J Med Genet B Neuropsychiatr Genet.

21. Rommelse, N.N., Arias-Vasquez, A., Altink, M.E., Buschgens, C.J., Fliers, E., Asherson, P., Faraone, S.V., Buitelaar, J.K., Sergeant, J.A., Oosterlaan, J., et al. (2008). Neuropsychological endophenotype approach to genome-wide linkage analysis identifies susceptibility loci for ADHD on 2q21.1 and 13q12.11. Am J Hum Genet 83, 99-105.

22. Zhou, K., Asherson, P., Sham, P., Franke, B., Anney, R.J., Buitelaar, J., Ebstein, R., Gill, M., Brookes, K., Buschgens, C., et al. (2008). Linkage to chromosome 1p36 for attention-deficit/hyperactivity disorder traits in school and home settings. Biol Psychiatry 64, 571-576.

23. Ogdie, M.N., Fisher, S.E., Yang, M., Ishii, J., Francks, C., Loo, S.K., Cantor, R.M., McCracken, J.T., McGough, J.J., Smalley, S.L., et al. (2004). Attention deficit hyperactivity disorder: fine mapping supports linkage to 5p13, 6q12, 16p13, and 17p11. Am J Hum Genet 75, 661-668.

24. Ogdie, M.N., Macphie, I.L., Minassian, S.L., Yang, M., Fisher, S.E., Francks, C., Cantor, R.M., McCracken, J.T., McGough, J.J., Nelson, S.F., et al. (2003). A genomewide scan for attention-deficit/hyperactivity disorder in an extended sample: suggestive linkage on 17p11. Am J Hum Genet 72, 1268-1279.

25. Arcos-Burgos, M., Castellanos, F.X., Pineda, D., Lopera, F., Palacio, J.D., Palacio, L.G., Rapoport, J.L., Berg, K., Bailey-Wilson, J.E., and Muenke, M. (2004). Attention-deficit/hyperactivity disorder in a population isolate: linkage to loci at 4q13.2, 5q33.3, 11q22, and 17p11. Am J Hum Genet 75, 998-1014.

26. Romanos, M., Freitag, C., Jacob, C., Craig, D.W., Dempfle, A., Nguyen, T.T., Halperin, R., Walitza, S., Renner, T.J., Seitz, C., et al. (2008). Genome-wide linkage analysis of ADHD using high-density SNP arrays: novel loci at 5q13.1 and 14q12. Mol Psychiatry 13, 522-530.

27. Neale, B.M., Lasky-Su, J., Anney, R., Franke, B., Zhou, K., Maller, J.B., Vasquez, A.A., Asherson, P., Chen, W., Banaschewski, T., et al. (2008). Genome-wide association scan of attention deficit hyperactivity disorder. Am J Med Genet B Neuropsychiatr Genet.

28. Aulchenko, Y.S., Vaessen, N., Heutink, P., Pullen, J., Snijders, P.J., Hofman, A., Sandkuijl, L.A., Houwing-Duistermaat, J.J., Edwards, M., Bennett, S., et al. (2003). A genome-wide search for genes involved in type 2 diabetes in a recently genetically isolated population from the Netherlands. Diabetes 52, 3001-3004.

29. Vaessen, N., Heutink, P., Houwing-Duistermaat, J.J., Snijders, P.J., Rademaker, T., Testers, L., Batstra, M.R., Sandkuijl, L.A., van Duijn, C.M., and Oostra, B.A. (2002). A genome-wide search for linkage-disequilibrium with type 1 diabetes in a recent genetically isolated population from the Netherlands. Diabetes 51, 856-859.

30. Pardo, L.M., MacKay, I., Oostra, B., van Duijn, C.M., and Aulchenko, Y.S. (2005). The effect of genetic drift in a young genetically isolated population. Ann Hum Genet 69, 288-295.

31. Aulchenko, Y.S., Heutink, P., Mackay, I., Bertoli-Avella, A.M., Pullen, J., Vaessen, N., Rademaker, T.A., Sandkuijl, L.A., Cardon, L., Oostra, B., et al. (2004). Linkage disequilibrium in young genetically isolated Dutch population. Eur J Hum Genet 12, 527-534.

32. Heutink, P., and Oostra, B.A. (2002). Gene finding in genetically isolated populations. Hum Mol Genet 11, 2507-2515.

33. Croes, E.A., El Galta, R., Houwing-Duistermaat, J.J., Ferdinand, R.F., Lopez Leon, S., Rademaker, T.A., Dekker, M.C., Oostra, B.A., Verhulst, F., and Van Duijn, C.M. (2005). Phenotypic subtypes in attention deficit hyperactivity disorder in an isolated population. Eur J Epidemiol 20, 789-794.

34. Ferdinand, R.F., and van der Ende, J. (2000). NIMH DISC-IV: Diagnostic Interview Schedule for Children. Authorized Dutch translation. In. (Erasmus MC-Sophia: Rotterdam, the Netherlands.

35. Shaffer, D., Fisher, P., Lucas, C.P., Dulcan, M.K., and Schwab-Stone, M.E. (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses. J Am Acad Child Adolesc Psychiatry 39, 28-38.

36. de Nijs, P.F., Ferdinand, R.F., de Bruin, E.I., Dekker, M.C., van Duijn, C.M., and Verhulst, D.C. (2004). Attention-deficit/hyperactivity disorder (ADHD): parents' judgment about school, teachers' judgment about home. Eur Child Adolesc Psychiatry 13, 315-320.

37. Miller, S.A., Dykes, D.D., and Polesky, H.F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16, 1215.

38. Liu, F., Kirichenko, A., Axenovich, T.I., Duijn, C.M.v., and Aulchenko, Y.S. (2008). An approach for cutting large and complex pedigrees for linkage analysis. European Journal of Human Genetics in press.

39. Aulchenko, Y.S., Bertoli-Avella, A.M., and van Duijn, C.M. (2005). A method for pooling alleles from different genotyping experiments. Ann Hum Genet 69, 233-238.

40. Mukhopadhyay, N., Almasy, L., Schroeder, M., Mulvihill, W.P., and Weeks, D.E. (1999). Mega2, a data-handling program for facilitating genetic linkage and association analyses. Am J Hum Genet 65, A436.

41. Lander, E.S., and Botstein, D. (1987). Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science 236, 1567-1570.

42. Liu, F., Elefante, S., van Duijn, C.M., and Aulchenko, Y.S. (2006). Ignoring distant genealogic loops leads to false-positives in homozygosity mapping. Ann Hum Genet 70, 965-970.

43. Abecasis, G.R., Cherny, S.S., Cookson, W.O., and Cardon, L.R. (2002). Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30, 97-101.

44. Freimer, N.B., Reus, V.I., Escamilla, M.A., McInnes, L.A., Spesny, M., Leon, P., Service, S.K., Smith, L.B., Silva, S., Rojas, E., et al. (1996). Genetic mapping using haplotype, association and linkage methods suggests a locus for severe bipolar disorder (BPI) at 18q22-q23. Nat Genet 12, 436-441.

45. Schumacher, J., Anthoni, H., Dahdouh, F., Konig, I.R., Hillmer, A.M., Kluck, N., Manthey, M., Plume, E., Warnke, A., Remschmidt, H., et al. (2006). Strong genetic evidence of DCDC2 as a susceptibility gene for dyslexia. Am J Hum Genet 78, 52-62.

46. Fisher, S.E., Marlow, A.J., Lamb, J., Maestrini, E., Williams, D.F., Richardson, A.J., Weeks, D.E., Stein, J.F., and Monaco, A.P. (1999). A quantitative-trait locus on chromosome 6p influences different aspects of developmental dyslexia. Am J Hum Genet 64, 146-156.

47. Gayan, J., Smith, S.D., Cherny, S.S., Cardon, L.R., Fulker, D.W., Brower, A.M., Olson, R.K., Pennington, B.F., and DeFries, J.C. (1999). Quantitative-trait locus for specific language and reading deficits on chromosome 6p. Am J Hum Genet 64, 157-164.

48. Grigorenko, E.L., Wood, F.B., Meyer, M.S., and Pauls, D.L. (2000). Chromosome 6p influences on different dyslexia-related cognitive processes: further confirmation. Am J Hum Genet 66, 715-723.

49. Willcutt, E.G., Pennington, B.F., Smith, S.D., Cardon, L.R., Gayan, J., Knopik, V.S., Olson, R.K., and DeFries, J.C. (2002). Quantitative trait locus for reading disability on chromosome 6p is pleiotropic for attention-deficit/hyperactivity disorder. Am J Med Genet 114, 260-268.

50. Brunner, D., Buhot, M.C., Hen, R., and Hofer, M. (1999). Anxiety, motor activation, and maternal-infant interactions in 5HT1B knockout mice. Behav Neurosci 113, 587-601.

51. Hawi, Z., Dring, M., Kirley, A., Foley, D., Kent, L., Craddock, N., Asherson, P., Curran, S., Gould, A., Richards, S., et al. (2002). Serotonergic system and attention deficit hyperactivity disorder (ADHD): a potential susceptibility locus at the 5-HT(1B) receptor gene in 273 nuclear families from a multi-centre sample. Mol Psychiatry 7, 718-725.

52. Quist, J.F., Barr, C.L., Schachar, R., Roberts, W., Malone, M., Tannock, R., Basile, V.S., Beitchman, J., and Kennedy, J.L. (2003). The serotonin 5-HT1B receptor gene and attention deficit hyperactivity disorder. Mol Psychiatry 8, 98-102.

Chapter 3

Genetic Studies of Personality Traits

Chapter 3.1

A genome wide linkage study of individuals with high scores on NEO personality traits

Najaf Amin[1], Maaike Schuur[1], Elena S. Gusareva[2], Aaron J. Isaacs[1], Yurii S. Aulchenko[1], Anatoly V. Kirichenko[3], Irina V. Zorkoltseva[3], Tatiana I. Axenovich[3,4], Ben A. Oostra[1,5], A. Cecile J.W. Janssens[1], Cornelia M. van Duijn[1*]

[1] Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands
[2] Department of Molecular and Cellular Immunology, Institute of Molecular Genetics AS CR, v.v.i. Prague, Czech Republic
[3] Institute of Cytology & Genetics, Russian Academy of Science, Novosibirsk, Russia
[4] Department of Cytology & Genetics, Novosibirsk State University, Novosibirsk, Russia
[5] Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, the Netherlands

ABSTRACT

The NEO Five Factor Inventory divides human personality traits into five dimensions: neuroticism, extraversion, openness, conscientiousness and agreeableness. In this study, we sought to identify genes with large effects for the five NEO personality traits by performing genome wide linkage analysis of individuals scoring in the extremes of these traits (> 90th percentile). Affected-only linkage analysis was performed using Illumina 6K linkage panel in a family-based study, the Erasmus Rucphen Family study (ERF). We subsequently determined whether distinct, segregating haplotypes found with linkage analysis were associated with the trait of interest in the population. Finally, a denser SNP genotyping array (Illumina 318K) was used to search for copy number variations (CNVs) in the associated regions. In the families with extreme phenotype scores, we found significant evidence of linkage for conscientiousness to 20p13 (rs1434789, LOD = 5.86) and suggestive evidence of linkage for neuroticism to 19q, 21q and 22q, extraversion to 1p, 1q, 9p and12q, openness to 12q, 19q, and agreeableness to 2p, 6q, 17q and 21q. Haplotype analysis showed unique haplotypes in 21q22 for neuroticism (p-values = 0.009, 0.007), in 17q24 for agreeableness (marginal p-value = 0.018) and in 20p13 for conscientiousness (marginal p-values = 0.058, 0.038) segregating in high LOD score contributing families. No evidence for CNVs in any of the associated regions was found. Our findings imply that there may be genes with large effects involved in personality traits, which may be identified with next generation sequencing techniques.

INTRODUCTION

Research into human personality has always been an integral part of behavioral science. Numerous epidemiological studies show that personality traits are important risk factors for many psychiatric and non-psychiatric disorders[1-9]. Personality traits are thought to be complex and determined largely by genetic factors[10]. Over the last few decades, various models for the assessment of personality have been developed, including the five factor model that divides human personality into five basic, universal types[11]. The NEO Five Factor Inventory (NEO-FFI) assesses these fundamental traits which include (1) neuroticism – a tendency to experience negative emotions, (2) extraversion – a predisposition towards enthusiasm, positive emotions and action, (3) openness – the extent of an individual's intellectual curiosity, (4) agreeableness – a propensity for cooperation and harmony, and (5) conscientiousness – the inclination to control, regulate and direct impulses[12]. Although the five scales are designed to be orthogonal, correlations appear due to self-reporting[13].

To date, genetic research has focused on neuroticism, primarily because of its association with disease and mortality[1-4; 6-9; 14-15]. Recent studies showed that other personality traits might also be important risk factors for different diseases; extraversion is a determinant of the future onset of bipolar disorder[16], conscientiousness is associated with Alzheimer disease (AD)[17], and several studies demonstrated that low scores in novelty seeking, a measure of creativity[18], are associated with a higher risk of developing Parkinson's disease[19-21]. There is increasing interest in the relationship between personality traits and treatment outcomes[22]. Another interesting aspect of personality is that these traits, specifically openness and conscientiousness, have been associated with higher academic achievement and better work performance[23].

Twin studies provided heritability estimates of the NEO personality traits ranging from 33 to 61%[10; 24-27]. Despite the high heritability, the genetics of these personality traits are not well understood. Candidate gene studies associated neuroticism with *SLC6A4*[28-30] and *TPH1*[31] and agreeableness, extraversion and conscientiousness with *ADH4* and *CHRM2*[32]. Genome wide association studies suggested association between the *CLOCK* gene and agreeableness and the *MAMDC1* and *NKAIN2* genes and neuroticism, however replication efforts were inconsistent[33-36]. There have been seven published genome-wide linkage scans for quantitative measures of personality traits, evaluated using either Eysenck's personality questionnaire (EPQ)[37] (measures neuroticism, extraversion and psychotism or the Tri-dimensional Personality Questionnaire (TPQ)[18; 38] (assesses novelty seeking, harm avoidance and reward dependence), as continuous outcomes including six exclusively for neuroticism[39-45]. These studies provided evidence for linkage between neuroticism/harm avoidance and several different loci (**Table 1**). Of the linked regions 1q, 8p, 11q, 12q and 18q were replicated for neuroticism in various studies[39-40; 42; 44].

Genome-wide association studies (GWAS) and linkage with quantitative outcomes target genes with small to moderate effects. We hypothesized that there may be genes with

large effects underlying the NEO personality traits and that these might be localized by studying persons scoring in the extremes for these phenotypes. We searched for such genes for all of the NEO-FFI personality traits by performing affected-only linkage analysis and haplotype association analysis. To translate the findings to the full distribution of the scale, we subsequently associated the haplotypes linked to the extremes to the outcomes on a quantitative scale.

MATERIAL AND METHODS

Study population

The study sample consisted of 2657 individuals who participated in the Erasmus Rucphen Family (ERF) study. The ERF cohort was ascertained from a genetically isolated region in the southwest of the Netherlands. The study population descended from 20 related couples that lived in the isolate between 1850 and 1900; genealogical data, available from 1750, captures all individuals in a single 23 generation pedigree including more than 23,000 individuals. Pedigree members 18 years and older were invited to participate. Spouses were invited only for descendents who had children older than 18 years.

Personality Assessment

The five basic personality types (neuroticism, extraversion, openness, agreeableness and conscientiousness) were assessed using the NEO Five Factor Inventory (NEO-FFI)[12]. The short form of this inventory consists of 60 items, 12 for each trait. Answers for each question were given on a 5-point scale and the total scores for each of the five traits ranged from 12 to 60. Because long questionnaires are often repetitious in nature and may cause the respondent to lose interest, the scores were considered invalid if the respondents gave the same response for at least 20 consecutive items (n = 114) or if they answered less than 9 questions in total (n = 176) on a single scale.

Genotyping

For all participants, genomic DNA was extracted from peripheral venous blood utilizing the salting out method[46]. For genome-wide linkage analysis, genotyping was performed using the Illumina 6K linkage panel that includes markers distributed evenly across the human genome (median inter-marker distance = 301 kb). We used 5250 single nucleotide polymorphisms (SNPs) after quality control (call rate > 95%) and the exclusion of X-chromosome SNPs. The genotyping was performed at the Centre National de Génotypage in France according to the manufacturer's protocol. For copy number variation analysis, we used an Illumina 318K SNP array. For this panel, micro-array based genotyping according to the manufacturer's instructions was performed at the Leiden Genome Technology Center of the Leiden University Medical Center.

Table 1: A summary of published genome wide linkage studies for neuroticism

Study	Questionnaire/trait	Chrom	Region (cM)	Markers	Lod score	P-value
Gillespie et al. 2008	JEPQ/Neuroticism	5	15	-	1.52[a]	0.008
		10	105-125	-	1.79[a]	0.004
		12	110	-	1.58[a]	0.007
		15	100	-	1.79[a]	0.004
		16	90-95	-	2.35[a]	0.001
		19	30	-	1.91[a]	0.003
Wray et al. 2008	EPQ/NEO/ Neuroticism	2	112	D2S1790	1.6	-
		5	191	D5S211	2.2	-
		6	75	D6S2410	1.5	-
		8	45	D8S1771	1.6	-
		8	134	D8S592	1.6	-
		10	5	D10S1412	2.0	-
		10	175	D10S212	1.7	-
		14	103	D14S1434	2.6	-
		15	17	GTTTT001	1.8	-
		18	117	D18S61	1.9	-
Kuo et. al. 2007	EPQ/Neuroticism	1	49	D1S470	1.77	-
		1	83	D1S476	1.38	-
		2	157	D2S349	1.07	-
		9	162	D9S1826	1.02	-
		11	43	D11S4080	2.06	-
		12	175	D12S1638	2.13	-
		15	124	D15S1014	1.03	-
		18	91	D18S68	1.39	-
		18	115	D18S1009	1.10	-
Neale et al. 2005	EPQ/Neuroticism	1	137	AMY	1.91[a]	0.003
		3	104.5	D3S2406	1.39[a]	0.0113
		6	147.3	D6S1003	1.39[a]	0.0112
		11	132	D11S4150	1.41[a]	0.0108
		12	45.5	D12S1042	2.22[a]	0.0014
Nash et al. 2004	EPQ/Neuroticism	1	80	D1S2890	1.6	0.62
		6	47	D6S1610	2.7	0.07
Fullerton et al. 2003	EPQ/Neuroticism	1	126	D1S2868	3.25[a]	0.00011
		4	176	D4S1539	3.15[a]	0.00014
		7	42	D7S516	3.18[a]	0.00013
		8	8	D8S277	2.29[a]	0.00117
		11	99	D11S898	3.00[a]	0.00020
		12	105	D12S346	3.95[a]	0.00002
		13	64	D13S153	3.12[a]	0.00015
Cloninger et al. 1998	TPQ/Harm-avoidance	8*	17	D8S1106	3.2	0.00006
		11	194	D11S1327	1.6	0.003
		18	109		1.6	0.004

EPQ = Eysenck personality questionnaire, JEPQ = Junior Eysenck personality questionnaire, TPQ = Tri-dimensional personality questionnaire. * region confirmed by Zohar et al. 2003 in an independent candidate region study

[a] LOD scores not reported in the actual studies but calculated for comparison in this study using the conversion χ^2/ 2*ln(10)

Table 2: General characteristics of the individuals selected for the linkage analysis

	Neuroticism	Extraversion	Openness	Agreeableness	Conscientiousness
Number	221	176	215	227	178
Age	47.5(14.2)	41.7(14.8)	42.1(13.5)	49.9(14.0)	50.0(14.0)
% women	70	54	61	78	56
Range in scale	12-59	16-60	13-52	20-60	23-60
Threshold score at 90th Percentile	42	48	40	51	54
Scores					
Neuroticism	45.9(3.8)	25.5(7.4)**	30.1(8.4)*	27.3(8.9)**	26.8(7.2)**
Extraversion	33.5(7.1)**	51.4(2.3)	42.3(5.9)**	42.8(6.9)**	45.1(6.5)**
Openness	32.4(6.1)*	34.5(6.8)**	43.8(2.7)	31.8(6.4)*	32.1(6.5)*
Agreeableness	42.4(6.4)**	46.5(5.9)**	43.8(6.3)	53.6(2.5)	47.1(6.6)**
Conscientiousness	42.3(6.9)**	51.6(5.4)**	46.1(7.0)	50.5(5.5)**	56.8(1.7)

* Difference between means of highest 10% and lowest 90% at $p < 0.05$, ** at $p < 0.001$
Values are means (standard deviation), unless otherwise indicated

Statistical Analyses

Linkage analysis

We performed affected only genome-wide linkage analysis by defining the affected as those people who scored above the 90th percentile for each of the five traits. The 90th percentile was chosen so as to maximize power based on our simulation studies evaluating different cut-offs[47]. The simulations suggest that for different sample size the power is highest for the 10% cut-off as published earlier[47]. General characteristics of the sample selected for the linkage analysis is given in **Table 2** and the correlation structure of these five traits at the population level is given in **Supplementary Table 1**.

The affected individuals for each trait were linked via a single large pedigree which was later clustered into small (≤ 18 bits) families for analytical purposes using pedigree clustering software[48]. We performed multipoint non-parametric (NPL) using the 'all' option in MERLIN[49], which uses Kong & Cox[50] linear model to evaluate the presence of linkage. We also performed parametric linkage analyses using both dominant and recessive models with complete penetrances and a disease allele frequency of 0.01 using MERLIN.The allele frequencies were estimated from all genotyped individuals in the pedigrees for each trait. On average (across all traits) 1249 Mendelian inconsistencies were observed for which we set the genotypes of the individuals within the whole family to missing.

Genome-wide simulations

Genome-wide significance thresholds were determined empirically by performing 500 genome-wide simulations under the null hypothesis of no linkage and using NPL with the 'all' option. Marker allele frequencies we set to the ones observed in our data. Pedigree clustering was performed using individuals having upper 10% of a value of a quantitative trait. For each genome screen, the highest log of odds (LOD) score was recorded. The cumulative density function of the simulated maximum LOD scores approximates the distribution of the genome-wide type I error rate. Our simulations showed that a LOD score of 4.1 corresponds to a genome-wide type I error rate of 5% and that an LOD of 2.8 corresponds to a genome-wide type I error of 50%. The NPL 'all' option was chosen for simulations since it would provide us with the most conservative threshold keeping in mind the structure of our pedigrees[51].

Haplotype construction & Association analysis

For regions showing significant or suggestive evidence of linkage, haplotypes were constructed for the families that contributed predominantly to the LOD score (contributing a LOD score of ≥ 1) in SIMWALK[52-53]. Haplotype association analysis was conducted using the "e" option in FBAT[54-55] to test for association in the presence of linkage, once the segregating haplotype was ascertained. The association analysis was run with an additive model when the highest LOD scores were observed with NPL. A recessive or a dominant model was used if the highest LOD score was obtained under one of those models. Also for binary trait haplotype association analysis (binary analysis) we used the highest scorers ($> 90^{th}$ percentile) as cases. In addition, we also performed a quantitative trait haplotype association analysis (quantitative analysis) in the total population. Frequencies of the haplotypes (*f*) were estimated from the total population. Bonferroni corrections were applied to the nominal p-values for the number of haplotypes tested per region. To get the estimate of odds ratio (OR) and effect size we also performed logistic and linear regression analyses in SPSS, for the haplotypes that showed evidence of association in FBAT, using the same model under which association was observed.

CNV analysis

In the regions that showed evidence of linkage and association, we searched for common copy number variants (CNVs) in individuals who were carrying the haplotype using the CNV partition tool in Beadstudio 2.4.4. We also used probe intensities (as measured by Log R Ratios) and genotype frequencies of the SNP probes (as measured by B-allele frequencies) to visually detect CNVs in the associated regions[56-57].

Table 3: Regions in suggestive or significant linkage with NEO-FFI traits

start SNP	end SNP	Cytogentic location	Position (start)	Position (end)	Dominant (HLOD)†	Recessive (HLOD)†	Non-parametric (LOD)∞	Genome-wide p-value*
Neuroticism								
rs964795	rs3499	19q13.4	63029177	63785296	2.50	*2.94*	*3.73*	0.11
rs1012959	rs1016694	21q22	36983492	38156688	0.15	0.51	*3.42*	0.21
rs713816	rs4444	22q11	24491639	29529888	2.60	*3.07*	2.50	0.37
Extraversion								
rs437749	rs1413527	1p31	82965819	88142058	0.14	0.43	*3.05*	0.38
rs913257	rs1004959	1q24-25	167253034	175983330	*2.88*	1.43	1.31	0.49
rs1532310	rs1412256	9p24	592986	1454067	0.43	0.10	*2.96*	0.44
rs1278602	rs7960480	12q24.3	132109288	132388516	1.94	2.06	*4.01*	0.06
Openness								
rs1278602	rs7960480	12q24.3	132109288	132388516	1.87	2.21	*3.71*	0.12
rs964795	rs3499	19q13.4	63029177	63785296	2.28	*3.00*	2.78	0.41
Agreeableness								
rs168293	rs6432244	2p25	8674900	12079927	0.51	1.43	*3.11*	0.36
rs727619	rs8770	6q27	170623826	170804163	1.69	0.86	*3.54*	0.16
rs411602	rs759563	17q24-25	66051172	67882483	0.92	0.63	*3.32*	0.25
rs2834380	rs2836301	21q22	34413370	38599459	1.99	0.80	*3.64*	0.14
Conscientiousness								
rs1434789	rs434609	20p13	85900	184992	**5.25**	**5.86**	**4.47**	0.03

Significant regions in bold
Suggestive regions in italic
† heterogeneity log of odds
∞ Kong & Cox log of odds
* Empirical genome-wide p-value estimated as a proportion of simulated scans resulting in NPL (LOD) ≥ the observed (500 simulations)

RESULTS

We report the Kong & Cox log of odds (LOD) score for the NPL and heterogeneity LOD score (HLOD) for the parametric linkage analyses. **Table 3** gives regions with significant and suggestive linkage in either of the parametric or the NPL analysis. **Table 4** provides the results of haplotype association analysis for the identified haplotypes including the ORs associated with the haplotypes.

Neuroticism

NPL analysis revealed two regions with suggestive evidence of linkage (LOD > 2.82) on chromosomes 19q13 and 21q22, and the recessive model gave one suggestive locus on chromosome 22q11 (**Figure 1A, Table 3**). None of the LOD scores reached genome-wide significance. We constructed haplotypes for the three suggestive regions (data not shown) in the families that were predominantly contributing to the LOD scores. For 21q22, we identified two unique haplotypes ($H21_{N1}$ & $H21_{N2}$ with frequencies of 0.47 and 0.26). $H21_{N1}$ was shared by six out of the nine individuals belonging to 3 high LOD score families (**Supplementary Figure 1**), five of which were homozygous carriers. Two of the remaining three were carriers of $H21_{N2}$. $H21_{N}$ covered 1.5 centiMorgan (cM) between rs2835574 and rs1016694. Both $H21_{N1}$ and $H21_{N2}$ showed nominally significant evidence of association in the binary analysis based on the total ERF population (p-value = 0.009, OR = 1.2 & p-value = 0.007, OR = 0.76). Both haplotypes remained significant after correction for multiple testing. $H21_{N}$ lies in the Down's syndrome critical region and covers 10 genes including *TTC3, DSCR9, DSCR3, DYRK1A, KCNJ6* and *PIGP* (**Table 4**).

Extraversion

Suggestive evidence of linkage was observed to chromosomes 1p31, 1q24, 9p24 and 12q24 in the NPL analysis (**Figure 1B, Table 3**). The LOD score for chromosome 12 (4.01) approached genome-wide significance. For 12q24, a distinct haplotype ($H12_{E}$) was observed in three high LOD score families (data not shown). $H12_{E}$ covered a 1 cM region between rs1278602 and rs7960480 and was shared by seven out of nine individuals with extreme phenotype belonging to the three families. Three of the seven were homozygous and 4 were heterozygous for the haplotype. $H12_{E}$ was the most frequent haplotype in the region ($f = 0.36$) but was not significant when tested for association (**Table 4**).

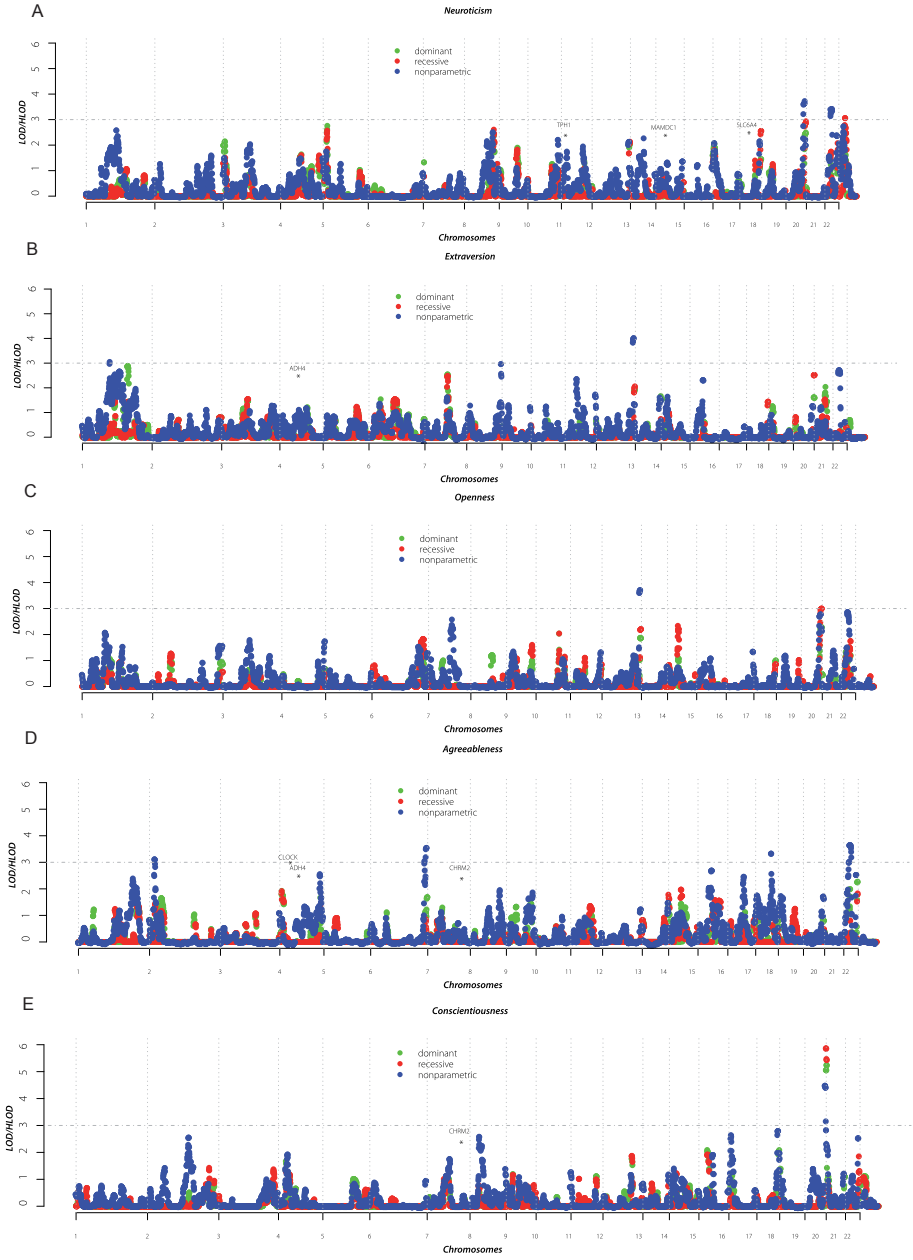


Figure 1. Genome wide linkage results from dominant (green), recessive (red) models, and non-parametric linkage (NPL) analysis (blue). The X-axis depicts the whole autosomal genome divided into 22 chromosomes. The Y-axis shows the heterogeneity LOD scores (HLOD) for dominant and recessive models and LOD scores from NPL analysis. (A) Neuroticism, (B) Extraversion, (C) Openness, (D) Agreeableness and (E) Conscientiousness.

Table 4. The association of the haplotypes identified in families showing evidence of linkage to the outcome of interest

Trait	chr	Haplotype Name	Haplotype Configuration	Flanking SNPs	Position	*f* [1]	p-value		OR (CI) Effect (SE)[2]	Genes
							Binary	quantitative		
Neuroticism										
	21	$H21_{N1}$	11111111111	rs2835574-rs1016694	37377194-38156688	0.473	0.009*	0.888	1.2(0.97,1.46)	TTC3, DSCR9,DSCR3, DYRK1A, KCNJ6
	21	$H21_{N2}$	22222222222	rs2835574-rs1016694	37377194-38156688	0.261	0.007*	0.520	0.76(0.57,0.99)	TTC3, DSCR9,DSCR3, DYRK1A, KCNJ6
Extraversion										
	12	$H12_{E}$	1111112	rs1278602-rs7960480	132109288-132388516	0.361	0.696	0.780		
Openness										
	12	$H12_{O}$	1121111	rs1278602-rs7960480	132109288-132388516	0.052	0.309	0.638		
Agreeableness										
	6	$H6_{A}$	12222	rs909472-rs8770	170731815-170804163	0.598	0.843	0.616		
	17	$H17_{A}$	111	rs411602-rs1981647	66051172-66132788	0.168	0.395	0.018*	0.31(0.23)	KCNJ2, KCNJ16, MAP2K6,
	21	$H21_{A1}$	222222222	rs2835574-rs2835653	37377194-37480506	0.521	0.366	0.943		
	21	$H21_{A2}$	111111111	rs2835574-rs2835653	37377194-37480506	0.255	0.349	0.689		
Conscientiousness										
	20	$H20_{C1}$	2221222	rs1434789-rs434609	85900-184992	0.231	0.058	0.152	1.4 (0.96,1.99)	DEFB, ZCCHC3, SOX12, NRSN2
	20	$H20_{C2}$	1112111	rs1434789-rs434609	85900-184992	0.339	0.408	0.094		

[1] frequency of the haplotype in the total population

Openness

For openness, there were two regions with a LOD > 2.82, one on chromosome 12q24, overlapping with extraversion, and one on 19q13, overlapping with neuroticism (**Figure 1C, Table 3**). This overlap is in line with the correlation structure of the traits (**Supplementary Table 1**) and the findings in **Table 2**. **Supplementary Table 1** shows a strong correlation in the overall population. Those scoring higher than the 90th percentile of the openness distribution were also significantly different in their means for neuroticism and extraversion when compared with individuals with lower scores on openness (**Table 2**). On 12q24, a unique haplotype ($H12_O$) was observed in 5 out of 14 individuals with extreme phenotypes from 4 contributing families (not shown). $H12_O$ covered a roughly 1 cM region between rs1278602 and rs7960480 and was relatively rare ($f = 0.05$) in the total ERF population. The haplotype differed from the one observed for extraversion in the same region. No evidence of association was observed for $H12_O$ in either binary or quantitative trait haplotype association analysis (**Table 4**).

Agreeableness

Suggestive evidence of linkage was observed on chromosomes 2p25, 6q27, 17q24, and 21q22 in NPL analysis (**Figure 1D, Table 3**). The 21q22 region was the same as that observed for neuroticism (**Table 3**), in accordance with the correlation structure (**Supplementary Table 1**) and the findings in Table 2 where those who scored high on agreeableness also scored low on neuroticism when compared with the less agreeable individuals (**Table 2**). In contrast to other NEO traits where there were only a few families predominantly contributing to the LOD score in the linked regions, for agreeableness there was no single family contributing a LOD score > 1. Instead there were multiple families each contributing positively to the total score. Construction of haplotypes in the three most contributing families revealed distinct haplotypes for chromosomes 6 ($H6_A$), 17 ($H17_A$) and 21 ($H21_A$) (not shown). Of these, only $H17_A$ (**Supplementary Figure 2**) showed significant evidence of association in the quantitative trait haplotype association analysis ($p = 0.018$, the change in agreeableness associated with carriership of 1 extra allele ($\beta = 0.31$). $H17_A$ had a frequency of 0.17, covers a 0.3 cM region and is flanked by rs411602 and rs1981647 (**Table 3**). The region covered by the haplotype did not contain any genes, but the nearby genes include *KCNJ2, KCNJ16* and *MAP2K6*.

Conscientiousness

A significant LOD score was observed for chromosome 20 (HLOD = 5.86, at rs1434789 under the recessive model, 5.28 under the dominant model and a LOD of 4.47 under NPL) (**Figure 1E, Table 3**). Construction of haplotypes in six families revealed two major segregating haplotypes ($H20_{c1}$ & $H20_{c2}$, **Supplementary Figure 3**) with frequencies of 0.339 and 0.231. Under a recessive model, $H20_{c2}$ showed borderline significance at a nominal level (p-value = 0.058, OR = 1.4) (**Table 4**). No association was observed under the dominant or additive models. The haplotype covers

a half cM region flanked by rs1434789 and rs434609.The region contains 5 genes, all belonging to the beta defensin protein family (DEFB). When tested for association under a recessive model, the top 10% scorers included only 8 families. To improve statistical power, we extended the association study to those scoring in the highest 30% of the distribution, enlarging the analysis to 25 families. In this analysis the haplotype became marginally significant (p-value = 0.038) but could not survive Bonferroni correction.

CNV analysis

Supplementary Figures 4, 5 and 6 show the CNV analysis of the regions 21q22, 17q24 and 20p13 for neuroticism, agreeableness and conscientiousness, respectively, for one carrier of $H21_{N1}$, ten carriers of $H17_{A}$ and four carriers of $H20_{C2}$. The CNV partition tool did not detect any CNV in the region. In the visual detection, the B-allele frequencies revealed a loss of heterozygosity (LOH) for homozygous carriers of the haplotypes in the regions, but the probe intensities did not point towards a deletion. Combining the two analyses makes it more likely that we are dealing with classical genomic variation for which individuals are homozygous by descent.

DISCUSSION

In our study of extended families, we found significant evidence (LOD > 4.12) of linkage for conscientiousness to 20p13 (rs1434789, LOD=5.86) and suggestive evidence of linkage for neuroticism to 19q, 21q and 22q, extraversion 1p, 1q, 9p and12q, openness to 12q, 19q, and agreeableness to 2p, 6q, 17q and 21q. Haplotype construction and association analysis showed unique haplotypes ($H21_{N1}$ & $H21_{N2}$) on 21q22 for neuroticism (p-values = 0.009, 0.007), on 17q24 for agreeableness (($H17_{A}$, p-value = 0.018) and on 20p13 for conscientiousness (($H20_{C2}$, p-value = 0.058) segregating in high LOD score contributing families. We did not detect evidence of CNVs in the regions; however, these analyses should be interpreted with caution given the low density of the array used.

In this study we aimed to find genes that are overrepresented in individuals with extreme phenotype. Although it has been argued that this may lead to a reduction of power as compared to a quantitative trait outcome in which all individuals are used[58], our simulation studies based on variants with large effect suggest that this a powerful approach[47]. When comparing the results discussed here to those derived from the analysis of the test results as quantitative outcomes (presented in **Supplementary Figure 7**), the overlap is limited. This is most likely explained by the fact that the extreme analysis will pick up genes with relative large effects. Yet at the 20p13 and 17q24 regions that show linkage to Conscientiousness and agreeableness respectively, there are linkage signals (albeit weak) also in the quantitative trait analysis.

Our study is the first to report linkage analysis of all 5 dimensions of the NEO personality traits. With haplotype construction and association analyses, we attempted to reduce the size of the regions and find segregating haplotypes that were associated with the traits. The construction of haplotypes also helped us reduce the size of the linked regions and thus avoid multiple testing issues in the subsequent association analyses to the extent that we only had to adjust for the number of haplotypes tested. Since we tested haplotypes only for a single trait, we did not adjust for the total number of haplotypes tested over the 5 scales. The family-based design also gave us an opportunity to look for CNVs that might be segregating in the population. It could be argued that utilizing a denser array than the 318K we used could have revealed smaller CNVs, in particular deletions. We observed some overlap of the linked regions between certain personality traits, for instance, the locus 19q13 was shared by neuroticism and openness, 21q22 by neuroticism and agreeableness, and 12q24 by extraversion and openness (**Table 3**). This is most likely due to the fact that strong correlations were observed between the personality traits (**Table 2 and supplementary Table 1**). These correlations, although theoretically unexpected for orthogonal scales, was also observed in previous studies[34]. It is interesting to note that most of our suggestive linkage peaks come from the NPL analysis but that these are usually followed by some what lower scores from both dominant and recessive models. This is in accordance with the work of Greenberg et al[59], which suggests that parametric LOD scores are more powerful compared to NPL only when the correct model (which is usually unknown for complex traits) is specified. Misspecification of parameters in the parametric linkage analysis has serious negative impact on its power to detect linkage[60].

When comparing the suggestive regions for neuroticism in our study with those of others, there was no overlap with previous linkage findings. This might be due to the fact that earlier studies used different instruments to assess neuroticism (or similar traits), such as Eysenck's Personality Questionnaire (EPQ) or the Tri-dimensional Personality Questionnaire (TPQ) (**Table 1**). Nevertheless, there were regions with moderate linkage signals that overlapped with those of previous studies. These include a locus on 5p, which gave a LOD score of 2.76 under a dominant model in our study (**Figure 1A**) for neuroticism, and overlaps with the finding of Gillespie et al.[41] (**Table 1**). Similarly, two loci on chromosome 11, 11p14 (rs1564745, LOD = 1.75) and 11q23 (rs1013582, LOD = 1.95), and one locus on chromosome 10 (rs6580, LOD = 1.26) overlap with previous findings on neuroticism (**Figure 1A, Table 1**)[39-40; 42; 44]. For extraversion, the 12q24 locus (LOD = 4.01)overlaps with one identified previously by Gillespie et al. (p-value = 0.009)[41]. This is, perhaps, the most interesting overlap, since with a LOD score of 4.01 this region almost fulfilled our strict threshold for significant linkage (LOD=4.1). From regions harboring studied candidate genes, mild linkage signals (LOD = 1.16) were observed for the TPH1 region for neuroticism and the ADH4 region for agreeableness (LOD = 1.33) and extraversion (LOD = 0.56) (**Figure 1B**). The overlap of our findings with those of others shows that our founder population in the southwestern part of the Netherlands is representative of the general population. This is in line

with the findings from our simulation studies that show that variants with frequencies > 0.001 are expected to be conserved in this population[61].

The most interesting region in our study was 20p13 for conscientiousness, which gave significant linkage signals under all three models. We were able to identify two haplotypes in this region segregating in high LOD score contributing families ($H21_{C1}$ & $H21_{C2}$). This haplotype covers a 100 kilobase (kb) region on 20p13 and contains 5 genes belonging to the beta defensin family (**Table 4**). Defensins play a role in the immunologic response to invading microorganisms. Other genes in the region include *ZCCHC3*, a gene with unknown function; *SOX12*, part of the SOX gene family that has been implicated in cell fate decisions in a diverse range of developmental processes involved is sex determination; and NRSN2, involved in the maintenance and/or transport of vesicles. This region was previously implicated in spinocerebellar ataxia[62] and neuro-degeneration[63]. $H20_{C2}$ showed only borderline significance (p-value = 0.058) under a recessive model in the haplotype association analysis despite the high LOD scores observed for this region (LOD = 5.86). An explanation for the borderline significance of $H20_{C2}$ may be the reduced number of families under the recessive model (8 families for the recessive model compared to 39 families for the additive model) for the estimation of the test statistic, resulting in a loss of power. This explanation was supported by the fact that a significant p-value (0.038) was observed when the analysis was run using the highest 30% as affected instead of the usual top10%. No significance was observed under additive or dominant models, which suggests a recessive mode of inheritance for this haplotype and its association with conscientiousness. Genes with recessive effects are difficult to identify in the general population but founder populations provide a strong vehicle to identify these genes.

The second most interesting region was 21q22 (LOD = 3.73) for neuroticism. Although the LOD score is not significant according to our stringent criteria, we found two haplotypes ($H21_{A1}$ & $H21_{A2}$) shared by individuals who had the extreme phenotypes. These haplotypes spread over 780 kb on 21q22 (**Table 3**), which is also known as the Down syndrome critical region. Association analysis showed significant results for $H21_{N1}$ (p-value = 0.009) and $H21_{N2}$ (p-value = 0.007) in the binary trait analysis. This region encompasses 10 genes, including *TTC3, DSCR9, DSCR3, DYRK1A, KCNJ6* and *PIGP*, and has been associated with cognitive and mental impairment in individuals with Down's syndrome[64-65]. *DYRK1A* has been associated with conscientiousness, which is strongly inversely correlated with neuroticism (see **Supplementary Table 1**)[34]. This is an interesting finding as many of the common physical and cognitive features of Down's syndrome may be present in apparently normal individuals.

The third most interesting locus was 17q24 for agreeableness (LOD = 3.32), since this region harbored a shared haplotype, $H17_{A,}$ that was observed in individuals with extreme phenotypes in high LOD score contributing families. $H17_{A}$ covered an 80 kb region on 17q24 and includes *KCNJ2, KCNJ16* and *MAP2K6* (**Table 3**). *KCNJ2* has been associated with Andersen syndrome, which is characterized by periodic paralysis, cardiac arrhythmias and dysmorphic

features[66] and *MAP2K6* has been associated with Alzheimer's disease[67]. In the association analysis $H17_A$ showed a p-value of 0.018 in the quantitative trait analysis but not the binary trait showing evidence of linkage. This may be explained by the fact that, unlike other personality traits, for agreeableness there was no single family contributing largely (LOD > 1) to the LOD scores. Instead, there were multiple families each contributing positively to the scores, which points towards the presence of a common variant rather than a rare one. Of interest is the fact that two of the three regions of interest harbor potassium/magnesium channels that are conserved in most mammalian cells (*KCNJ2, KCNJ6, KCNJ16*) while in the other region one of the genes (*SOX12*) is known to interact with a potassium channel (*KCNJ10*).

Other interesting regions include 19q13 and 22q11 for neuroticism. Chromosome 19q13 is a very important region that contains the apolipoprotein E gene (*APOE*). Both neuroticism and *APOE* are known to be associated with AD[1; 5; 68]. The 22q11 region is of interest since it was found to be associated with schizophrenia[69], while neuroticism is a known risk factor of schizophrenia[9; 16]. Finally the 12q24 region for extraversion, which showed borderline significance in our study, was found earlier by Gillespie et al.[41] in a linkage analysis of the same trait. Additionally, this region was previously associated with schizophrenia and bipolar disorder[70-71].

In this study, we attempted to identify genes with relatively large effects on personality traits. Although the estimates of odds ratio and effect size are moderate, these effects are rare compared the low ORs seen in GWAS (1.01 < OR < 1.15). GWAS may still yield novel genes with small effects with large enough samples, there is growing interest in whole genome sequencing aiming to indentify loci with large effects. Our study suggests that such loci can be identified for personality traits. Our linkage analysis revealed three interesting regions (20p13 for conscientiousness, 21q22 for neuroticism and 17q24-25 for agreeableness), which may include genes that exert moderate to large effects on NEO personality traits. The overlap with previous linkage studies, and with associated personality disorders, suggests that the findings in our population can be extrapolated to other populations. We confirm linkage of conscientiousness with 20p13, of neuroticism with 21q22 and of agreeableness with 17q24. The other identified regions need to be followed up with deep sequencing, which may help in capturing the actual mutation responsible for the linkage signals.

***Acknowledgements**:*

The ERF study was further supported by grants from the Netherlands Organisation for Scientific Research (NWO), Erasmus university medical center, the Centre for Medical Systems Biology (CMSB1 and CMSB2) of the Netherlands Genomics Initiative (NGI). We are grateful to all patients and their relatives, general practitioners and neurologists for their contributions and to Petra Veraart for her help in genealogy, Jeannette Vergeer for the supervision of the laboratory work and Peter Snijders for his help in data collection. ESG was supported by The International Human Frontier Science Program Organization.

REFERENCES

1. Crowe, M., Andel, R., Pedersen, N.L., Fratiglioni, L., and Gatz, M. (2006). Personality and risk of cognitive impairment 25 years later. Psychol Aging 21, 573-580.
2. Goodwin, R.D., and Stein, M.B. (2003). Peptic ulcer disease and neuroticism in the United States adult population. Psychother Psychosom 72, 10-15.
3. Hettema, J.M., Neale, M.C., Myers, J.M., Prescott, C.A., and Kendler, K.S. (2006). A population-based twin study of the relationship between neuroticism and internalizing disorders. Am J Psychiatry 163, 857-864.
4. Kendler, K.S., Gatz, M., Gardner, C.O., and Pedersen, N.L. (2006). Personality and major depression: a Swedish longitudinal, population-based twin study. Arch Gen Psychiatry 63, 1113-1120.
5. Meins, W., and Dammast, J. (2000). Do personality traits predict the occurrence of Alzheimer's disease? Int J Geriatr Psychiatry 15, 120-124.
6. Robertson, D.A., Ray, J., Diamond, I., and Edwards, J.G. (1989). Personality profile and affective state of patients with inflammatory bowel disease. Gut 30, 623-626.
7. Shipley, B.A., Weiss, A., Der, G., Taylor, M.D., and Deary, I.J. (2007). Neuroticism, extraversion, and mortality in the UK Health and Lifestyle Survey: a 21-year prospective cohort study. Psychosom Med 69, 923-931.
8. Terracciano, A., Lockenhoff, C.E., Crum, R.M., Bienvenu, O.J., and Costa, P.T., Jr. (2008). Five-Factor Model personality profiles of drug users. BMC Psychiatry 8, 22.
9. Van Os, J., and Jones, P.B. (2001). Neuroticism as a risk factor for schizophrenia. Psychol Med 31, 1129-1134.
10. Jang, K.L., Livesley, W.J., and Vernon, P.A. (1996). Heritability of the big five personality dimensions and their facets: a twin study. J Pers 64, 577-591.
11. McCrae, R.R. (2002). NEO-PI-R datafrom 36 cultures: Further intercultural comparisons. In The five-factor model of personality across cultures, R.R. McCrae and A. J, eds. (New York, Kluwer Academic/Plenum), pp 105-125.
12. Costa, P.T., and McCrae, R.R. (1992). Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI) Professional Manual. Psychological Assessment Resources.
13. Biesanz, J.C., and West, S.G. (2004). Towards understanding assessments of the big five: multitrait-multimethod analyses of convergent and discriminant validity across measurement occasion and type of observer. J Pers 72, 845-876.
14. Winstow, D.J., Wakefield, J.A., and Goldsmith, W.M. (1990). The relationship between personality, health symptoms and disease. Person individ Diff 11, 717-723.
15. Zbogar, and Hema. (2004). Neuroticism as a potential risk factor for Alzheimer's disease. The Journal of Addiction and Mental Health.
16. Lonnqvist, J.E., Verkasalo, M., Haukka, J., Nyman, K., Tiihonen, J., Laaksonen, I., Leskinen, J., Lonnqvist, J., and Henriksson, M. (2009). Premorbid personality factors in schizophrenia and bipolar disorder: results from a large cohort study of male conscripts. J Abnorm Psychol 118, 418-423.
17. Wilson, R.S., Schneider, J.A., Arnold, S.E., Bienias, J.L., and Bennett, D.A. (2007). Conscientiousness and the incidence of Alzheimer disease and mild cognitive impairment. Arch Gen Psychiatry 64, 1204-1212.
18. Cloninger, C.R., Svrakic, D.M., and Przybeck, T.R. (1993). A psychobiological model of temperament and character. Arch Gen Psychiatry 50, 975-990.
19. Kaasinen, V., Nurmi, E., Bergman, J., Eskola, O., Solin, O., Sonninen, P., and Rinne, J.O. (2001). Personality traits and brain dopaminergic function in Parkinson's disease. Proc Natl Acad Sci U S A 98, 13272-13277.
20. Menza, M.A., Golbe, L.I., Cody, R.A., and Forman, N.E. (1993). Dopamine-related personality traits in Parkinson's disease. Neurology 43, 505-508.
21. Menza, M. (2000). The personality associated with Parkinson's disease. Curr Psychiatry Rep 2, 421-426.
22. Canuto, A., Giannakopoulos, P., Meiler-Mititelu, C., Delaloye, C., Herrmann, F.R., and Weber, K. (2009). Personality traits influence clinical outcome in day hospital-treated elderly depressed patients. Am J Geriatr Psychiatry 17, 335-343.

23. Poropat, A.E. (2009). A meta-analysis of the five-factor model of personality and academic performance. Psychol Bull 135, 322-338.

24. Waller, N. (1999). Evaluating the structure of personality. In Personality and psychopathology, C. Cloninger, ed. (Washington, D.C., American Psychiatric Association), pp 155-197.

25. Riemann, R., Angleitner, A., and Strelau, J. (1997). Genetic and environmental influences on personality: A study of twins reared together using the self- and peer report NEO-FFI scales. J Personal 65, 449-475.

26. Loehlin, J.C., McCrae, R.R., Paul T. Costa, j., and John, O.P. (1998). Heritabilities of common and measure-specific components of the big five personality factors. Journal of research in personality 32, 431-453.

27. Jang, K.L., McCrae, R.R., Angleitner, A., Riemann, R., and Livesley, W.J. (1998). Heritability of facet-level traits in a cross-cultural twin sample: support for a hierarchical model of personality. J Pers Soc Psychol 74, 1556-1565.

28. Lesch, K.P., Bengel, D., Heils, A., Sabol, S.Z., Greenberg, B.D., Petri, S., Benjamin, J., Muller, C.R., Hamer, D.H., and Murphy, D.L. (1996). Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. Science 274, 1527-1531.

29. Mazzanti, C.M., Lappalainen, J., Long, J.C., Bengel, D., Naukkarinen, H., Eggert, M., Virkkunen, M., Linnoila, M., and Goldman, D. (1998). Role of the serotonin transporter promoter polymorphism in anxiety-related traits. Arch Gen Psychiatry 55, 936-940.

30. Sen, S., Burmeister, M., and Ghosh, D. (2004). Meta-analysis of the association between a serotonin transporter promoter polymorphism (5-HTTLPR) and anxiety-related personality traits. Am J Med Genet 127B, 85-89.

31. Nash, M.W., Sugden, K., Huezo-Diaz, P., Williamson, R., Sterne, A., Purcell, S., Sham, P.C., and Craig, I.W. (2005). Association analysis of monoamine genes with measures of depression and anxiety in a selected community sample of siblings. Am J Med Genet B Neuropsychiatr Genet 135B, 33-37.

32. Luo, X., Kranzler, H.R., Zuo, L., Zhang, H., Wang, S., and Gelernter, J. (2007). CHRM2 variation predisposes to personality traits of agreeableness and conscientiousness. Hum Mol Genet 16, 1557-1568.

33. Hettema, J.M., van den Oord, E.J., An, S.S., Kendler, K.S., and Chen, X. (2009). Follow-up association study of novel neuroticism gene MAMDC1. Psychiatr Genet.

34. Terracciano, A., Sanna, S., Uda, M., Deiana, B., Usala, G., Busonero, F., Maschio, A., Scally, M., Patriciu, N., Chen, W.M., et al. (2008). Genome-wide association scan for five major dimensions of personality. Mol Psychiatry.

35. van den Oord, E.J., Kuo, P.H., Hartmann, A.M., Webb, B.T., Moller, H.J., Hettema, J.M., Giegling, I., Bukszar, J., and Rujescu, D. (2008). Genomewide association analysis followed by a replication study implicates a novel candidate gene for neuroticism. Arch Gen Psychiatry 65, 1062-1071.

36. Calboli, F.C., Tozzi, F., Galwey, N.W., Antoniades, A., Mooser, V., Preisig, M., Vollenweider, P., Waterworth, D., Waeber, G., Johnson, M.R., et al. (2010). A genome-wide association study of neuroticism in a population-based sample. PLoS One 5, e11504.

37. Eysenck, S.B.G., Eysenck, H.J., and Barrett, P. (1985). A revised version of the psychotism scale. Person individ Diff 6, 21-29.

38. Cloninger, C.R., Przybeck, T.R., and Svrakic, D.M. (1991). The Tridimensional Personality Questionnaire: U.S. normative data. Psychol Rep 69, 1047-1057.

39. Cloninger, C.R., Van Eerdewegh, P., Goate, A., Edenberg, H.J., Blangero, J., Hesselbrock, V., Reich, T., Nurnberger, J., Jr., Schuckit, M., Porjesz, B., et al. (1998). Anxiety proneness linked to epistatic loci in genome scan of human personality traits. Am J Med Genet 81, 313-317.

40. Fullerton, J., Cubin, M., Tiwari, H., Wang, C., Bomhra, A., Davidson, S., Miller, S., Fairburn, C., Goodwin, G., Neale, M.C., et al. (2003). Linkage analysis of extremely discordant and concordant sibling pairs identifies quantitative-trait loci that influence variation in the human personality trait neuroticism. Am J Hum Genet 72, 879-890.

41. Gillespie, N.A., Zhu, G., Evans, D.M., Medland, S.E., Wright, M.J., and Martin, N.G. (2008). A genome-wide scan for Eysenckian personality dimensions in adolescent twin sibships: psychoticism, extraversion, neuroticism, and lie. J Pers 76, 1415-1446.

42. Kuo, P.H., Neale, M.C., Riley, B.P., Patterson, D.G., Walsh, D., Prescott, C.A., and Kendler, K.S. (2007). A genome-wide linkage analysis for the personality trait neuroticism in the Irish affected sib-pair study of alcohol dependence. Am J Med Genet B Neuropsychiatr Genet 144B, 463-468.

43. Nash, M.W., Huezo-Diaz, P., Williamson, R.J., Sterne, A., Purcell, S., Hoda, F., Cherny, S.S., Abecasis, G.R., Prince, M., Gray, J.A., et al. (2004). Genome-wide linkage analysis of a composite index of neuroticism and mood-related scales in extreme selected sibships. Hum Mol Genet 13, 2173-2182.

44. Neale, B.M., Sullivan, P.F., and Kendler, K.S. (2005). A genome scan of neuroticism in nicotine dependent smokers. Am J Med Genet B Neuropsychiatr Genet 132, 65-69.

45. Wray, N.R., Middeldorp, C.M., Birley, A.J., Gordon, S.D., Sullivan, P.F., Visscher, P.M., Nyholt, D.R., Willemsen, G., de Geus, E.J., Slagboom, P.E., et al. (2008). Genome-wide linkage analysis of multiple measures of neuroticism of 2 large cohorts from Australia and the Netherlands. Arch Gen Psychiatry 65, 649-658.

46. Miller, S.A., Dykes, D.D., and Polesky, H.F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16, 1215.

47. Van Gestel, S., Houwing-Duistermaat, J.J., Adolfsson, R., van Duijn, C.M., and Van Broeckhoven, C. (2000). Power of selective genotyping in genetic association analyses of quantitative traits. Behav Genet 30, 141-146.

48. Liu, F., Kirichenko, A., Axenovich, T.I., Duijn, C.M.v., and Aulchenko, Y.S. (2008). An approach for cutting large and complex pedigrees for linkage analysis. European Journal of Human Genetics in press.

49. Abecasis, G.R., Cherny, S.S., Cookson, W.O., and Cardon, L.R. (2002). Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30, 97-101.

50. Kong, A., and Cox, N.J. (1997). Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 61, 1179-1188.

51. McPeek, M.S. (1999). Optimal allele-sharing statistics for genetic mapping using affected relatives. Genet Epidemiol 16, 225-249.

52. Sobel, E., and Lange, K. (1996). Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. Am J Hum Genet 58, 1323-1337.

53. Sobel, E., Papp, J.C., and Lange, K. (2002). Detection and integration of genotyping errors in statistical genetics. Am J Hum Genet 70, 496-508.

54. Horvath, S., Xu, X., and Laird, N.M. (2001). The family based association test method: strategies for studying general genotype--phenotype associations. Eur J Hum Genet 9, 301-306.

55. Laird, N.M., Horvath, S., and Xu, X. (2000). Implementing a unified approach to family-based tests of association. Genet Epidemiol 19 Suppl 1, S36-42.

56. Conlin, L.K., Thiel, B.D., Bonnemann, C.G., Medne, L., Ernst, L.M., Zackai, E.H., Deardorff, M.A., Krantz, I.D., Hakonarson, H., and Spinner, N.B. Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis. Hum Mol Genet.

57. Yau, C., and Holmes, C.C. (2008). CNV discovery using SNP genotyping arrays. Cytogenet Genome Res 123, 307-312.

58. Almasy, L., and Blangero, J. (1998). Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet 62, 1198-1211.

59. Durner, M., Vieland, V.J., and Greenberg, D.A. (1999). Further evidence for the increased power of LOD scores compared with nonparametric methods. Am J Hum Genet 64, 281-289.

60. Clerget-Darpoux, F., Bonaiti-Pellie, C., and Hochez, J. (1986). Effects of misspecifying genetic parameters in lod score analysis. Biometrics 42, 393-399.

61. Pardo, L.M., MacKay, I., Oostra, B., van Duijn, C.M., and Aulchenko, Y.S. (2005). The effect of genetic drift in a young genetically isolated population. Ann Hum Genet 69, 288-295.

62. Verbeek, D.S., van de Warrenburg, B.P., Wesseling, P., Pearson, P.L., Kremer, H.P., and Sinke, R.J. (2004). Mapping of the SCA23 locus involved in autosomal dominant cerebellar ataxia to chromosome region 20p13-12.3. Brain 127, 2551-2557.

63. Zhou, B., Westaway, S.K., Levinson, B., Johnson, M.A., Gitschier, J., and Hayflick, S.J. (2001). A novel pantothenate kinase gene (PANK2) is defective in Hallervorden-Spatz syndrome. Nat Genet 28, 345-349.

64. Altafaj, X., Dierssen, M., Baamonde, C., Marti, E., Visa, J., Guimera, J., Oset, M., Gonzalez, J.R., Florez, J., Fillat, C., et al. (2001). Neurodevelopmental delay, motor abnormalities and cognitive deficits in transgenic mice overexpressing Dyrk1A (minibrain), a murine model of Down's syndrome. Hum Mol Genet 10, 1915-1923.

65. Capone, G.T. (2001). Down syndrome: advances in molecular biology and the neurosciences. J Dev Behav Pediatr 22, 40-59.

66. Andelfinger, G., Tapper, A.R., Welch, R.C., Vanoye, C.G., George, A.L., Jr., and Benson, D.W. (2002). KCNJ2 mutation results in Andersen syndrome with sex-specific cardiac and skeletal muscle phenotypes. Am J Hum Genet 71, 663-668.

67. Zhu, X., Rottkamp, C.A., Hartzler, A., Sun, Z., Takeda, A., Boux, H., Shimohama, S., Perry, G., and Smith, M.A. (2001). Activation of MKK6, an upstream activator of p38, in Alzheimer's disease. Journal of neurochemistry 79, 311-318.

68. van der Flier, W.M., Schoonenboom, S.N., Pijnenburg, Y.A., Fox, N.C., and Scheltens, P. (2006). The effect of APOE genotype on clinical phenotype in Alzheimer disease. Neurology 67, 526-527.

69. Karayiorgou, M., Morris, M.A., Morrow, B., Shprintzen, R.J., Goldberg, R., Borrow, J., Gos, A., Nestadt, G., Wolyniec, P.S., Lasseter, V.K., et al. (1995). Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. Proc Natl Acad Sci U S A 92, 7612-7616.

70. Reif, A., Herterich, S., Strobel, A., Ehlis, A.C., Saur, D., Jacob, C.P., Wienker, T., Topner, T., Fritzen, S., Walter, U., et al. (2006). A neuronal nitric oxide synthase (NOS-I) haplotype associated with schizophrenia modifies prefrontal cortex function. Mol Psychiatry 11, 286-300.

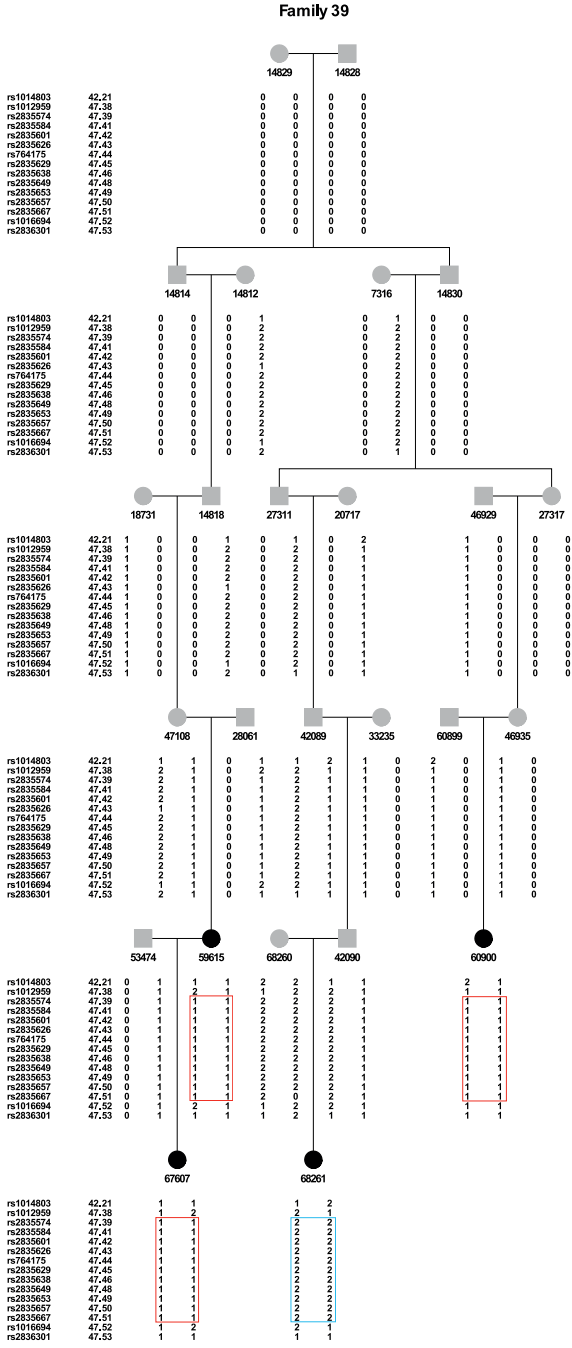
71. Reif, A., Strobel, A., Jacob, C.P., Herterich, S., Freitag, C.M., Topner, T., Mossner, R., Fritzen, S., Schmitt, A., and Lesch, K.P. (2006). A NOS-III haplotype that includes functional polymorphisms is associated with bipolar disorder. The international journal of neuropsychopharmacology / official scientific journal of the Collegium Internationale Neuropsychopharmacologicum (CINP) 9, 13-20.

SUPPLEMENTARY INFORMATION

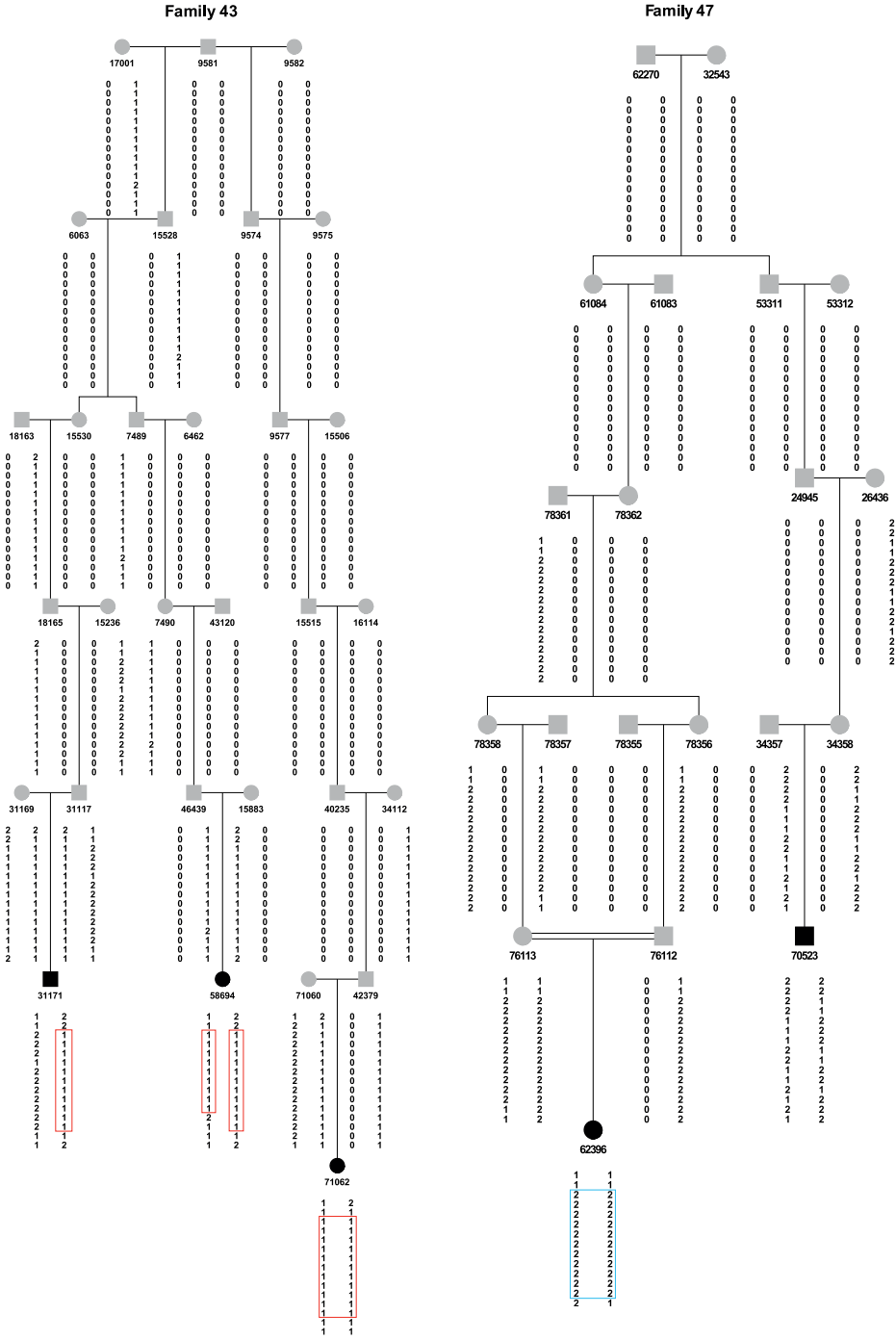
Supplementary table 1: Partial correlation coefficients for all the NEO personality traits in the general population adjusted for age and sex

Traits		Neuro-ticism	Extra-version	Open-ness	Agree-ableness	Conscien-tiousness
Neuroticism	Correlation	1.000	-.470	-.081	-.304	-.353
	P-value	NA	.000	.000	.000	.000
Extraversion	Correlation	-.470	1.000	.145	.268	.445
	P-value	.000	NA	.000	.000	.000
Openness	Correlation	-.081	.145	1.000	-.049	-.056
	P-value	.000	.000	NA	.017	.007
Agreeableness	Correlation	-.304	.268	-.049	1.000	.311
	P-value	.000	.000	.017	NA	.000
Conscientiousness	Correlation	-.353	.445	-.056	.311	1.000
	P-value	.000	.000	.007	.000	NA

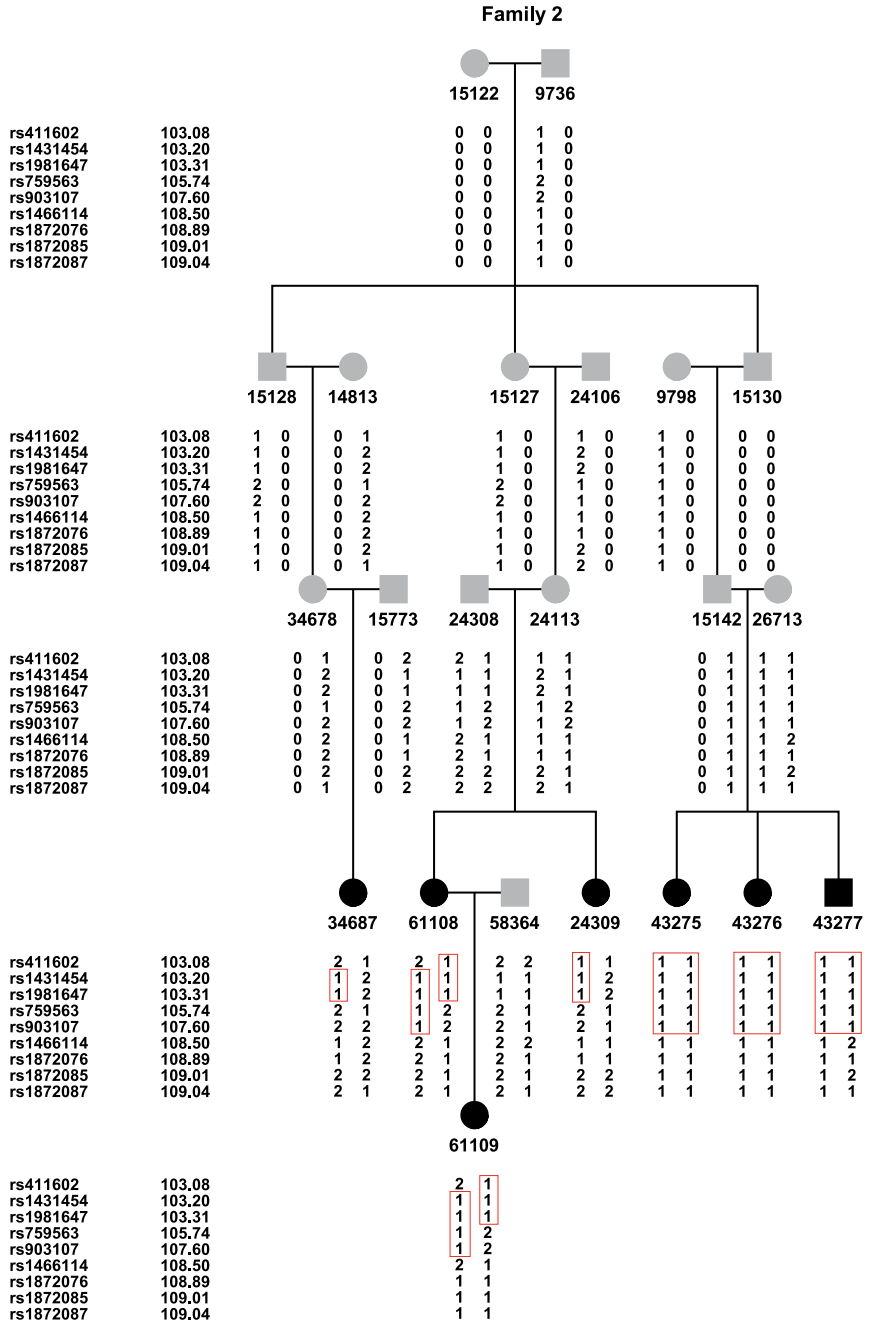
Supplementary Figure 1. Three high LOD score contributing families for neuroticism for the chromosome 21q22. Males are depicted with squares and females with circles. Solid squares are affected males and solid circles are affected females. Two distinct shared haplotypes are marked by red and blue boxes.



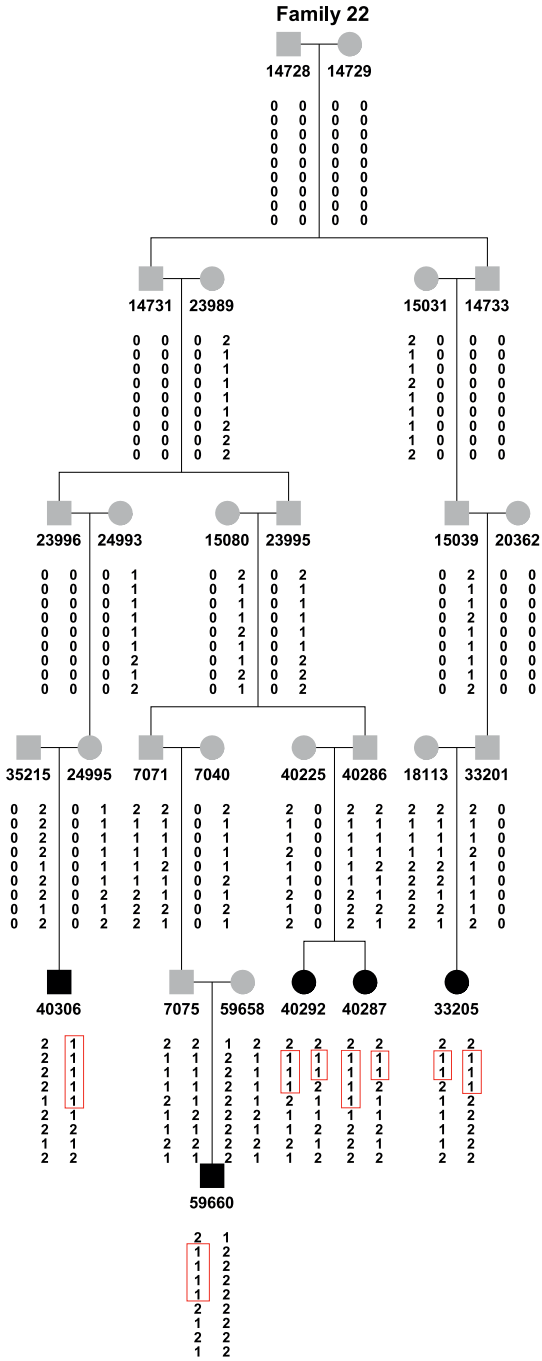
Supplementary Figure 1. (Continued)



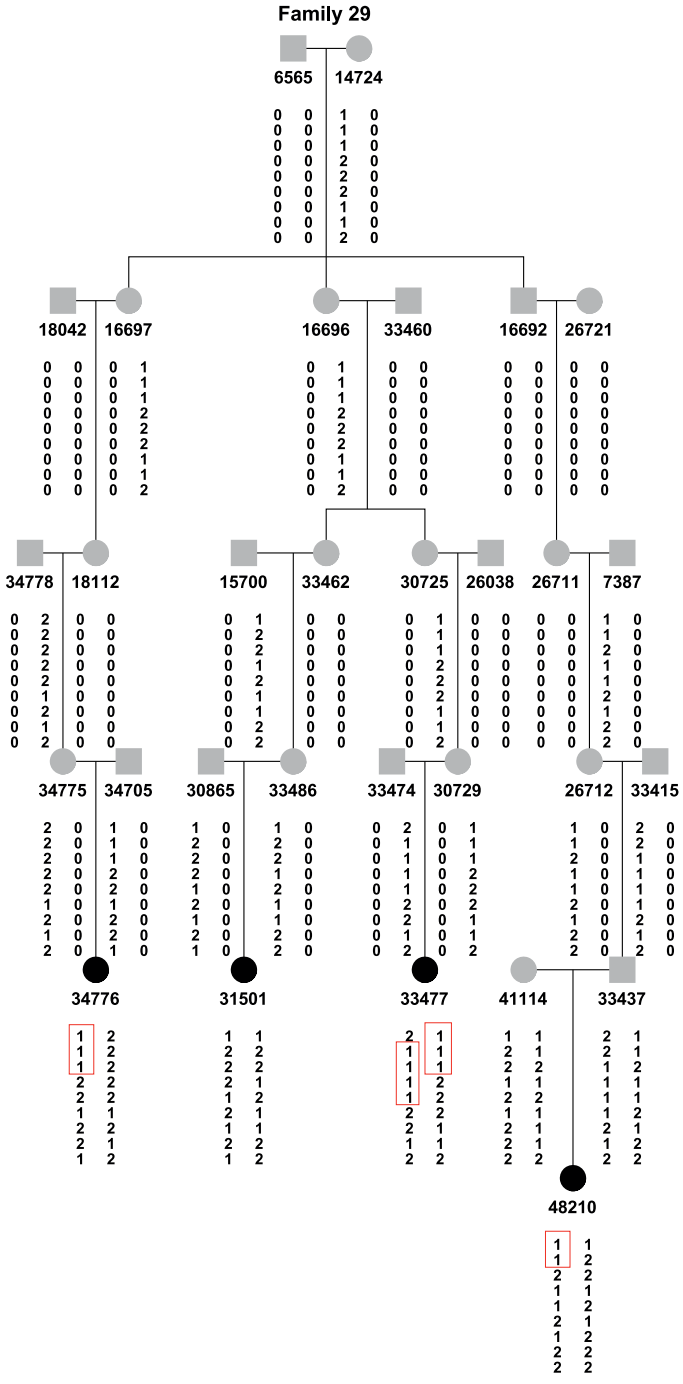
Supplementary Figure 2. Three high LOD score contributing families for agreeableness for the chromosome 17q24-25. Males are depicted with squares and females with circles. Solid squares are affected males and solid circles are affected females. Shared haplotype is marked by a red box.



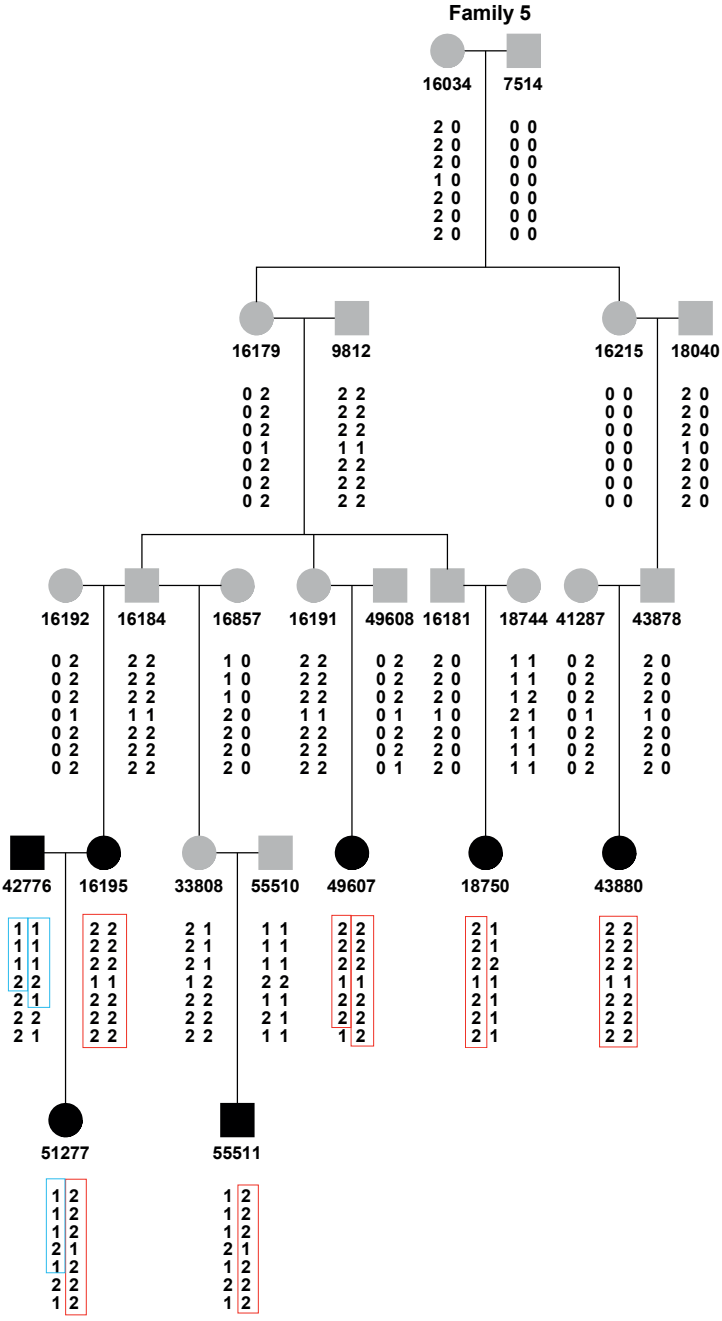
Supplementary Figure 2. (Continued)



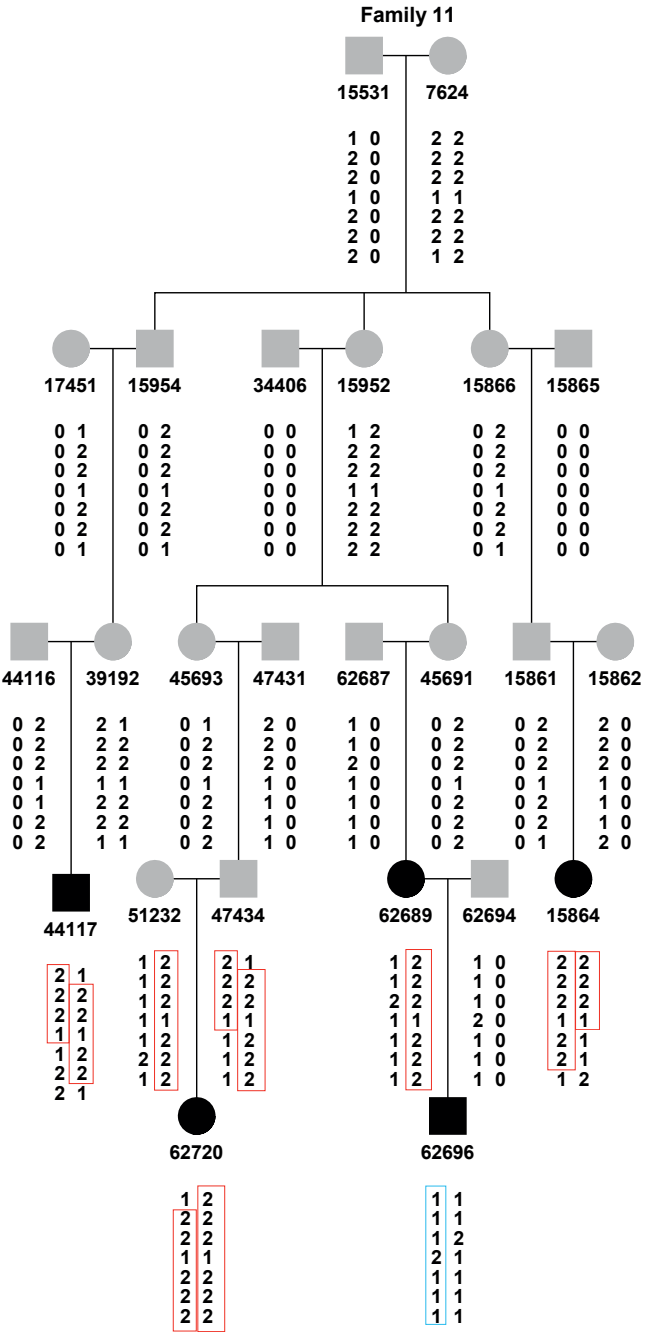
Supplementary Figure 2. (Continued)



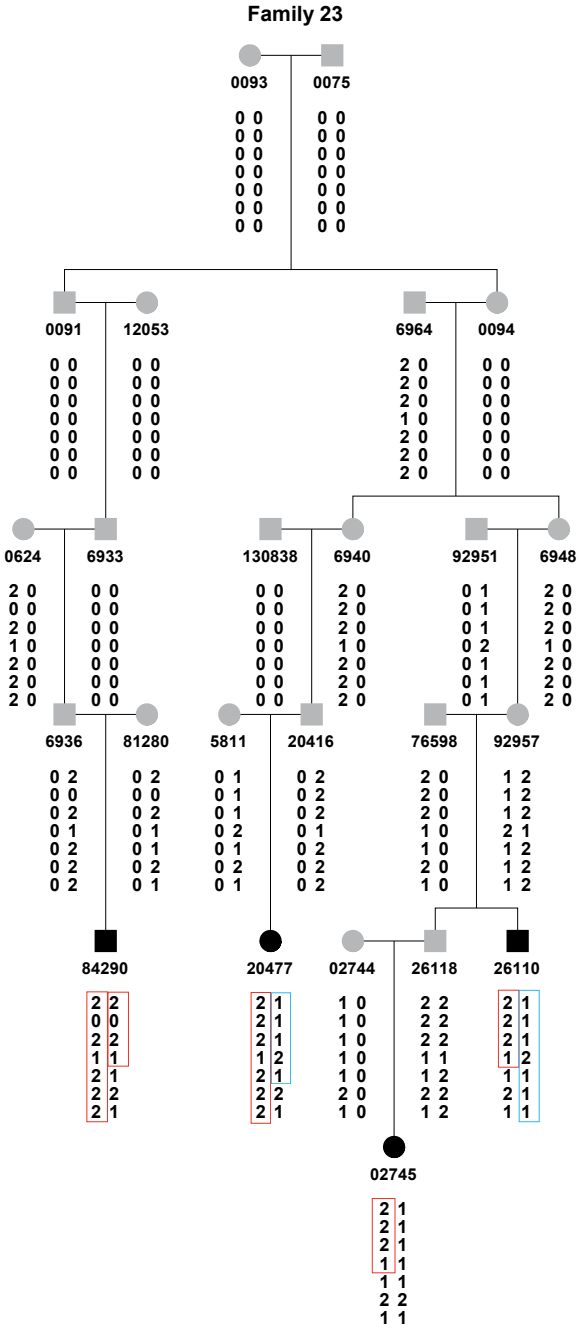
Supplementary Figure 3. Three high LOD score contributing families for conscientiouness for the chromosome 20p13. Males are depicted with squares and females with circles. Solid squares are affected males and solid circles are affected females. Two distinct shared haplotypes are marked by red and blue boxes.



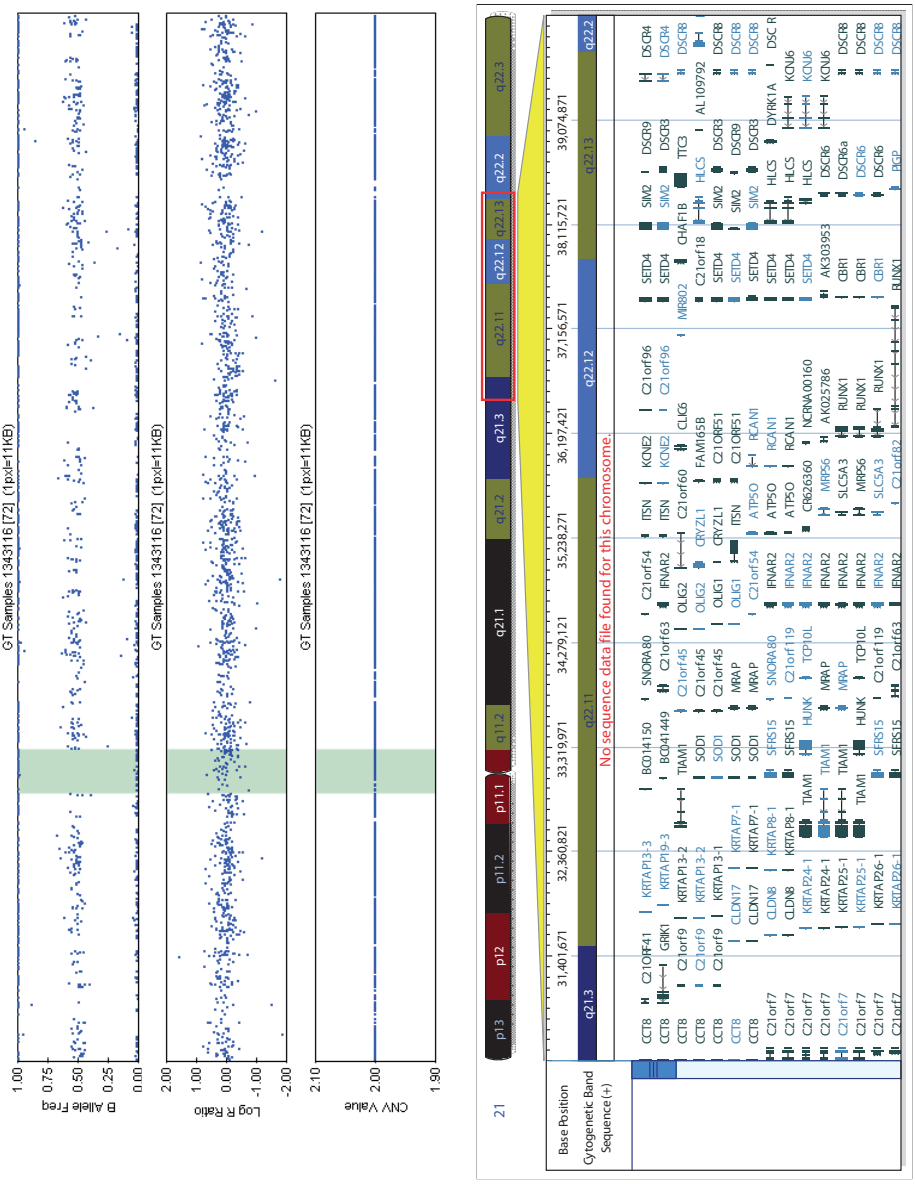
Supplementary Figure 3. (Continued)



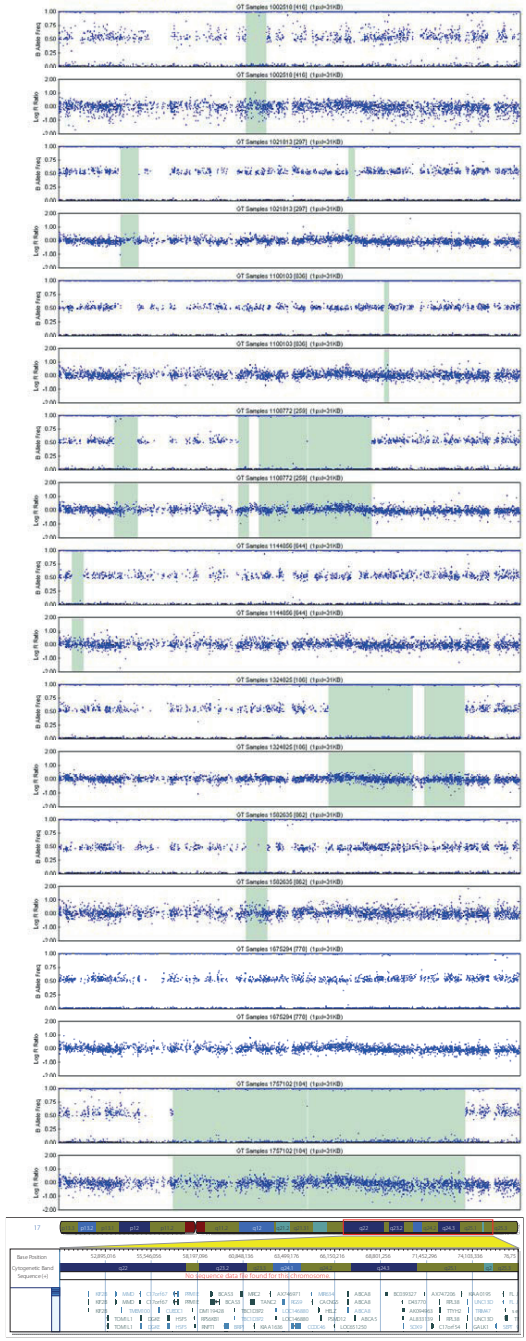
Supplementary Figure 3. (Continued)



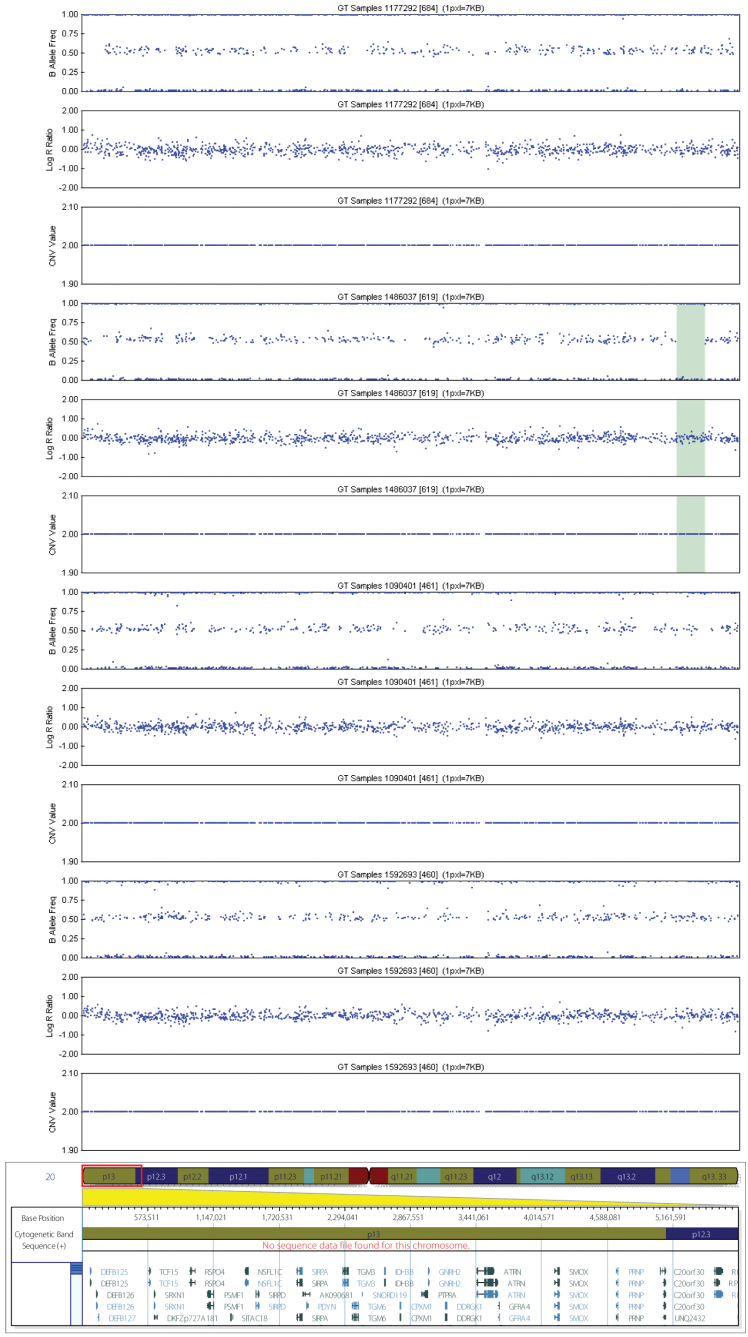
Supplementary Figure 4. CNV analysis for one highly neurotic individual who was also a homozygote carrier of H21N1. The top panel shows the frequency of the SNP probes (B-allele frequency) and the lower panel shows the probe intensities (Log R ratio) in the region of the associated haplotype H21N1. Region with excess homozygosity is highlighted.



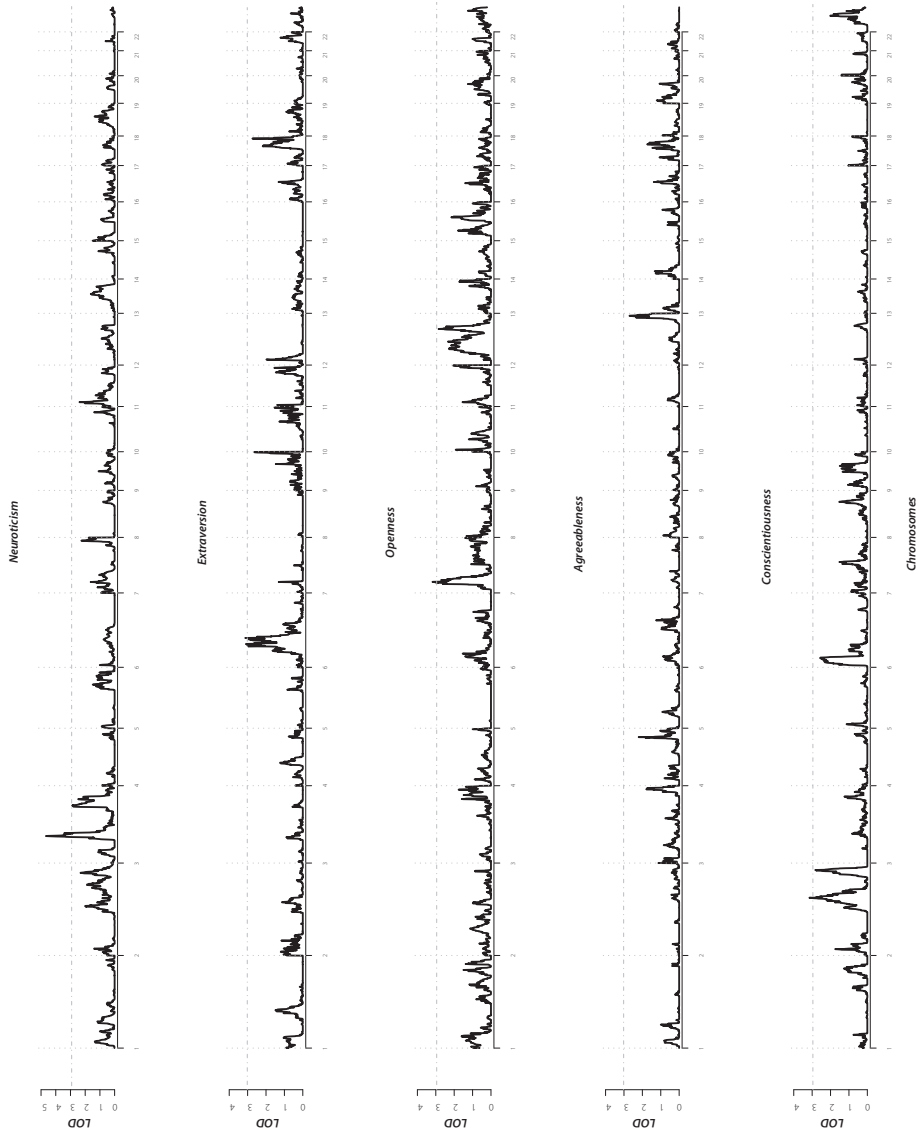
Supplementary Figure 5. CNV analysis for 10 highly agreeable individuals. The top panel shows the frequency of the SNP probes (B-allele frequency) and the lower panel shows the probe intensities (Log R ratio) in the region of the associated haplotype H17A. Regions with excess homozygosity are highlighted.



Supplementary Figure 6. CNV analysis for 4 extremely conscientious individuals. The top panel shows the frequency of the SNP probes (B-allele frequency) and the lower panel shows the probe intensities (Log R ratio) in the region of the associated haplotype H20C1.



Supplementary Figure 7. Results of quantitative trait linkage analysis for each of the five personality traits performed using variance components linkage analysis in MERLIN. From top to bottom, neuroticism, extraversion, openness, agreeableness and conscientiousness. The horizontal axis shows the whole autosomal genome divided into 22 chromosomes. The vertical axis shows the LOD scores.



Chapter 3.2

Refining genome-wide linkage intervals using GWAS identifies several loci influencing NEO Personality Dimensions

Najaf Amin,[1]* Viatcheslav Saviouk,[2]* Narelle K. Hansell,[3]* A. Cecile J.W. Janssens,[4] Marleen H.M. de Moor,[2] Pamela A. F. Madden,[5] Irina V. Zorkoltseva,[6] Brenda W. Penninx,[7,8,9] Antonio Terracciano,[10] Manuela Uda,[11] Toshiko Tanaka,[10] Tonu Esko,[12] Anu Realo,[12] Luigi Ferrucci,[10] Michelle Luciano,[13] Gail Davies,[13] Andres Metspalu,[12] Goncalo R. Abecasis,[14] Ian J. Deary,[13] Katri Raikkonen,[15] Laura J. Bierut,[5] Paul T. Costa,[10] Jouke J. Hottenga,[2] Gu Zhu,[3] Anatoly V. Kirichenko,[6] Aaron J. Isaacs,[1] Yurii S. Aulchenko,[1] Gonneke Willemsen,[2] Andrew C. Heath,[5] and Michele L Pergadia,[5] Sarah E. Medland,[3] Tatiana I. Axenovich,[6] Eco de Geus,[2] Grant W. Montgomery,[3] Margaret J. Wright,[3] Ben A. Oostra,[1,16] Nicholas G. Martin,[3]* Dorret I. Boomsma,[2]* Cornelia M. van Duijn[1,17]*†

[1]Unit of Genetic Epidemiology, Department of Epidemiology, Erasmus University Medical Center, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands, [2]Department of Biological Psychology, VU University Amsterdam, 1081 BT, Amsterdam, The Netherlands, [3]Queensland Institute of Medical Research, Brisbane, QLD, Australia, [4]Department of Epidemiology, Erasmus University Medical Center, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands, [5]Department of Psychiatry, Washington University School of Medicine, 660 S. Euclid, CB 8134, St. Louis, MO 63108, USA, [6]Institute of Cytology & Genetics, Russian Academy of Science, Novosibirsk, Russia [7]Department of Psychiatry, University Medical Center Groningen, Groningen, The Netherlands [8]Departments of Clinical Psychology and Psychiatry, Leiden University, Leiden, The Netherlands [9]Department of Psychiatry, EMGO+ Institute, Neuroscience Campus Amsterdam, VU University Medical Center Amsterdam, Amsterdam, The Netherlands, [10]National Institute on Aging, NIH, Baltimore, MD, USA, [11]Istituto di Neurogenetica e Neurofarmacologia, CNR, Monserrato, Cagliari, Italy, [12]Estonian Genome Project, University of Tartu and Estonian Biocentre, Tartu, Estonia, [13]Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, The University of Edinburgh, 7 George Square, Edinburgh, EH8 9JZ, UK, [14]Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA, [15]Department of Psychology, University of Helsinki, P.O. Box 9 (Siltavuorenpenger 20 D), 00014 University of Helsinki, Finland, [16]Department of clinical Genetics, Erasmus University Medical Center, the Netherlands, [17]Centre of Medical Systems Biology, Netherlands Consortium on Health Aging and National Genomics Initiative

* Authors contributed equally

ABSTRACT

Personality traits are complex phenotypes related to social and health outcomes, including psychiatric and somatic disorders. Individually, various gene finding methods have not achieved much success in finding genetic variants that account for differences in personality traits. We performed a meta-analysis of 4 unpublished genome-wide linkage scans (*N* = 6149 subjects) of five basic personality traits assessed with the NEO-FFI. We compared the significant regions to the results of a recent meta-analysis of genome-wide association studies (GWAS) (*N* ~ 17,000). We found significant evidence for linkage for neuroticism to chromosome 3p14 (rs1490265, LOD = 4.67) and to chromosome 19q13 (rs628604, LOD = 3.55), for extraversion to 14q32 (ATGG002, LOD = 3.3); for agreeableness to 3p25 (rs709160, LOD = 3.67) and to two adjacent regions on chromosome 15; 15q13 (rs970408, LOD = 4.07) and 15q14 (rs1055356, LOD = 3.52) in the individual scans. In the meta-analysis, we found strong evidence of linkage of extraversion to 9q34, 10q24 and 11q22, openness to 2p25, 3q26, 9p21, 11q24, 15q26 and 19q13, and agreeableness to 4q34 and 19p13. When combining these data with the association results of the GWAS of these personality traits, significant evidence of association was detected for openness at 11q24 (rs677035, p-value = $2.6*10^{-06}$). Borderline evidence for association was detected between neuroticism and rs332389 (p-value = $9*10^{-05}$) at 3p14 and between extraversion and rs7088779 (p-value = $4.2*10^{-06}$) at 10q24. Our study suggests that combining linkage and association is a useful approach in the identification of genes involved in complex phenotypes such as personality.

INTRODUCTION

Human personality is a compound of complex traits that are associated with several psychiatric[1-16], and somatic disorders[17-26]. Despite high heritability ranging from 33% to 60%, the understanding of the genetic origins of personality trait variation is extremely limited. Where linkage analyses have identified several large regions with suggestive linkage, very few overlap. Candidate gene studies have their own concerns in that the findings have generally not been replicated (see chapter 3.1). A recent genome-wide association study (GWAS) including up to 18,000 individuals has yielded only a few loci that attained genome-wide significance[27].

The general success of GWAS in gene discovery and failure to replicate most of the significant linkage peaks for complex traits and diseases has shifted interest towards GWAS. Association studies have benefitted from precise estimation of the locus, reproducibility of the findings and the availability of large population-based cohorts, which has lead to powerful studies for gene discovery as well as replication using meta-analyses. However, there is increasing debate on the question whether GWAS in the long run will be able to unravel the genetics of complex disease completely. The 'case of missing heritability[28] has led to the view that common traits may be driven by variants with frequencies between 0.001 and 0.01, which are not included in the present arrays and cannot be imputed reliably, although other explanations have also been suggested[29]. It is hypothesized that these variants would be unveiled through sequencing the entire genome. As it will be difficult to find and replicate rare variants in the general population, family based studies, which are believed to be enriched for rare variants are expected to provide a powerful platform for the discovery of such variants using various sequencing techniques. However, the return of family studies also raises the question whether meta-analyses of family-based studies will work. Because of financial restrictions, there are few large families subject to whole genome/ whole exome sequencing to date. We have evaluated this question therefore, using linkage analyses of large family-based studies and addressing personality as a quantitative outcome. Until large scale sequencing becomes inexpensive and various statistical tools are developed to enhance power for detecting the presence of rare variants in huge datasets, combining genome-wide linkage and association studies might be useful in discovering such variants. With this two pronged strategy linkage can overcome the problem of locus heterogeneity and association can compensate for the failure of linkage to pinpoint the actual genes.

Despite the fact that rare variants are weakly tagged by the common single nucleotide polymorphisms (SNPs) present in the current arrays[29-30], association signals in GWAS at loci including rare variants have been seen in particular for lipids[28]. Also two recent papers argued that common variants may tag more rare ones from a theoretical perspective[29-30] implying that linkage and association analyses should be able to identify the same loci. For complex traits it was suggested to use genome-wide linkage and association analyses methods in a two step

analysis to maximize power[31].

In this study we have combined the information from the results of four unpublished genome-wide linkage scans of personality traits (neuroticism, extraversion, openness, agreeableness and conscientiousness) and their meta-analysis with the results of a recently published meta-analysis of GWAS of the same traits[27]. This is by far the largest linkage study conducted for the NEO personality traits combined with the largest GWAS of these traits conducted to date.

MATERIALS AND METHODS

We performed a meta-analysis of linkage studies that used the NEO Five Factor Inventory (NEO-FFI) or the Revised NEO personality inventory (NEO-PI-R)[32] to assess the five basic personality traits including: (1) neuroticism—a trait that refers to the tendency to experience negative emotions: (2) extraversion – a measure of sociability, positive emotions and action: (3) openness — a measure of intellectual curiosity and preference for variety: (4) agreeableness — a measure of altruism, cooperation and harmony,: and (5) conscientiousness — a measure of an individual's tendency to plan, organize and direct his impulses[32]. This five-factor model is a hierarchical model where each of these five traits is defined by six underlying facets assessed by the NEO-PI-R. These five traits, also known as the Big Five, are considered universal[33] and stable in adulthood[34]. The five factors are expected to be orthogonal, but correlations appear possibly due to self-report [35]. Women score generally higher on Neuroticism and Agreeableness[36-38] compared to men. The NEO-FFI consists of 60 items, 12 for each trait, whereas the NEO-PI-R consists of 240 questions inclusive of the ones in the NEO-FFI. Answers are given on these on a 5-point scale, which are then added to give a total score. The current study included unpublished linkage scans from four independent populations including the Erasmus Rucphen Family (ERF), the Netherlands Twin Register (NTR), the Australian Adolescent sample (QIMR_adolescent) and the Australian adult sample (QIMR_adult) (**Table1**). ERF and NTR used the NEO-FFI for personality assessment whereas the Australian samples were assessed with either NEO-FFI or NEO-PI-R but for this study the 60 items of the NEO-FFI was used for the final assessment. Within each cohort the scores were considered invalid if an individual answered fewer than nine questions otherwise the missing data were imputed by taking the individual's average score for the valid items of that dimension.

Table1: General Features of independent genome scan

Cohort	ERF	VU	QIMR_Adolescents	QIMR_Adults
Origin	Dutch	Dutch	Australian*	Australian*
Design	Extended families	sibpairs	sibpairs	Sibpairs+Extended families
Sample size (%women)	2244 (55.0)	1507 (61.6)	1096(56.5)	1349(55.7)
Total # Markers used	5250	864	5479	5479
Marker Type	Single neucleotide polymorphisms	Microsatellites	Single neucleotide polymorphisms	Single neucleotide polymorphisms
Analysis Type	Variance components	Variance component	Variance components	Variance components
Analysis Software	Merlin	Merlin	Merlin	Merlin
Personality assessment	NEO-FFI	NEO-FFI	NEO-PI-R/ NEO-FFI	NEO-PI

* Majority of sample is Causasian, predominantly Ango-Celtic (ancestry outliers, identified using HapMap3 and GenomeEUTwin individuals as a reference panel, were excluded).

Erasmus Rucphen Family (ERF)

The study sample consisted of 2657 individuals who participated in Erasmus Rucphen Family (ERF) study[39]. The study population essentially consists of one extended family spanning over 23 generations and including more than 23,000 individuals descending from 20 related couples that lived in the region between 1850 and 1900. All descendants were ascertained and descendants of 18 years and older were invited to participate. Spouses were invited only for family members that had children of 18 years and older.

For all participants genomic DNA was extracted from peripheral venous blood utilizing the salting out method[40]. For genome-wide linkage analysis genotyping was performed using the Illumina 6K linkage panel that includes markers distributed evenly across the human genome (median distance between markers is 301 kb). Of the 6000 single nucleotide polymorphisms (SNPs) we used 5250, after quality control and excluding X-chromosomal SNPs. The genotyping was performed at the Centre National de Génotypage in France according to the standard protocol.

Variance component multipoint linkage analysis was performed using MERLIN v.1.0.1 software[41] for all the five personality traits. The pedigrees were split on non-overlapping fragments of no more than 18 bits with the help of two programs: Jenti[42] and PedSTR[43]. Three sets of sub-pedigrees were obtained with different parameters with the help of these programs, which were then analyzed separately. These three sets differed one from another by number, size and structure of pedigree fragments. However, they demonstrated similar profiles of LODs for all analyzed traits (Spearman correlations varied from 0.6 (p-value < 0.001) to 0.8 (p-value <

0.001), which allowed us to use the maximum of three values of LOD for each marker locus[44]. In accordance with Bonferroni correction suggestive and significant thresholds were estimated as 2.34 and 3.75, respectively.

The analysis was based on 2244 genotyped and phenotyped persons from ERF. All traits were adjusted on sex and age. Marker allele frequencies were estimated from the data.

Netherlands Twin Register (NTR)

The Netherlands Twin Register (NTR)[45] focuses on longitudinal phenotypic and biological data collection in Dutch twins and their family members. NTR linkage sample consists of 711 families with 3,412 non-clone individuals (1,438 founders, 1,870 females) with average 4.8 subjects per family. 282 of these families have both founders genotyped and 138 families had one genotyped founder. In addition, there are 290 nuclear families with no genotyped founders, and one extended pedigree with four founders without genotypes. Autosomal genomes had 757 markers spaced at an average of 4.76 cM (range 0.0-20.59 cM), with average heterozygosity of 0.76. Founders had the genetic data for 446 autosomal microsatellites. NEO measures together with the age and gender information were available for a total of 998 non-founder subjects (383 males; 615 females) from 409 families with genetic data. Of 409 families, 270 had two phenotyped siblings, 113 had three, 19 had four, one family had five, four families had six, and two families had seven phenotyped siblings resulting in total of 835 sibling pairs. In addition, this sample also features 83 phenotyped MZ clones.

The genetic maps were obtained through the Rutgers University Map Interpolator[46]. The allele frequencies were estimated with the Mendel v.10.0 software package[41]. VC linkage scan of the autosomal genome was conducted with Merlin v.1.1.2[32] using the multipoint identity-by-descent (IBD) information with age and sex as covariates at 1 cM resolution.

Australian Study Sample (QIMR)

NEO personality data (NEO-PI-R or NEO-FFI) were collected as part of two independent research streams – one focused on an adult sample (QIMR_adult) and the other with an adolescent/young adult focus (QIMR_adolescent). The QIMR_adult data were collected as part of the Nicotine Addiction Genetics (NAG) study (2001-2005), which targeted families based on heavy smoking index cases identified in earlier interviews and questionnaires[47-48-49], and was itself part of the Interactive Research Project Grants (IRPG). This sample comprised 1349 genotyped individuals aged 21 to 85 years (M = 45.5±13.1), from 519 families, and included 15 complete MZ pairs for whom data were averaged.

The QIMR_adolescent data were collected from two population studies under the umbrella of the Brisbane Adolescent and Young Adult Twin Studies, specifically, studies of cognition (1996-ongoing)[50] and health and well-being (2002-2003)[50], and from a study of borderline personality disorder (2004-2006)[51]. This sample comprised 1096 genotyped individuals aged

16 to 27 years (M = 19.4±2.7), from 563 families, and included 127 complete MZ pairs for whom data were averaged.

Participants were typically Caucasian, predominantly Ango-Celtic (ancestry outliers, identified using HapMap3 and GenomeEUTwin individuals as a reference panel, were excluded). Written, informed consent was obtained from all participants and from a parent or guardian for those aged under 18. Ethics approval was received from the institutional review boards appropriate to each study (Queensland Institute of Medical Research and Washington University School of Medicine).

SNP selection for linkage analysis was matched as far as possible to the SNP set used with the ERF sample. The final selection contained 5479 SNPs, of which 5181 had a direct match with our genotyped or imputed data. For the remaining 298 SNPs, proxies based on linkage disequilibrium (> 0.8), or position, were used.

Genotyping was done using the Illumina 610K or 370K SNP platform and Illumina BeadStudio software, with 269,840 SNPs common to the subsamples passing QC (28% of the SNPs selected for linkage were from this set of SNPs). Data were imputed using HapMap I+II (CEU, build 36, release 22) and MACH, as described in Medland et al.[52]. Variance component multipoint linkage analyses of the five personality domains were run for the adolescent and adult samples using MERLIN 1.1.2[41]. Sex and age were included as covariates in all analyses.

Meta-analysis of linkage scans

Results from individual genome-wide linkage scans were combined together in a meta-analysis using the Genome Search Meta-Analysis (GSMA) method[53] as implemented in the GSMA program[54]. The GSMA method divides the genome into several bins of equal width (chosen arbitrarily). The criteria of selecting bin-width includes that the bins are not too narrow so that correlation arises between adjacent bins owing to the wide linkage peaks usually extending to 30 centiMorgans (cM); and not too wide so the smallest chromosome should have at least 2 bins. Traditionally a bin width of 30 cM, which gives 120 bins across the autosomal genome, is used. The highest log of odds (LOD) score in each bin is then recorded and the bins are ranked in descending order in each sample with the bin with the highest LOD score getting the highest rank. The ranks within a bin are summed across each study to get a summed rank (SR). Bins with higher SR indicate evidence of linkage across the studies. The SR statistic is tested for significance using its distribution function or by simulation[54] which gives the probability of observing a given SR for a bin (P_{SR}). P_{SR} only gives the point-wise probability for the SR for a certain bin. A genome-wide interpretation of the results is obtained through the ordered statistic (OR), which determines the probability (P_{OR}) of a given SR for a bin by chance when bins are assigned ranks randomly in multiple simulations [55]. Simulations show that a bin with a significant P_{SR} and a significant P_{OR} ($P_{SR} < 0.05$ & $P_{OR} < 0.05$) has a high probability of containing a true susceptibility locus[55]. For an individual bin the genome-wide significance is defined as $P_{SR} < 0.05$/number of

bins and suggestive as P_{SR} < 1/number of bins. For an individual scan we used 3.3 as significant and 1.9 as suggestive linkage threshold.

In this study except for the NTR sample genome-wide linkage results were reported against SNP markers which helped us map all the results directly to the base pair positions. For NTR we mapped all the results to base pair positions by interpolation and using base pair positions from Rutgers map [46]. For each study we divided the autosomal genome into 125 bins of width 25 million base pairs (mbp) each. This division gave a maximum of 10 bins on chromosome 1 and a minimum of 2 bins on chromosome 22. To evaluate the possibility of correlation between adjacent bins, we performed the analysis using a bin-width of 40mbp also. We performed weighted meta-analysis, where weights were calculated as the square root of the sample size in each study. 10,000 permutations were performed to get the P_{SR} and P_{OR}. A bin was considered significant if (1) either the Bonferroni corrected significance was achieved i.e. P_{SR} < 0.0004 for a bin or (2) if both P_{SR} and P_{OR} were nominally significant for multiple bins [55]. Suggestive linkage threshold was set at P_{SR} < 0.008. We also tested for heterogeneity between studies using the HEGSMA [56] software.

In an attempt to discover the variant that might be causing the linkage signals, we fine mapped the regions with significant evidence of linkage in each individual study and significant and suggestive bins of the meta-analysis with the results of a meta-analysis of genome-wide association studies (GWAS) of the same NEO personality traits[27]. The meta-analysis of GWAS (n > 17000) included the samples used in this study (and many others as well) but the marker sets used were different.

Table 2: Significant and suggestive linkage regions in individual linkage scans

Study/trait	Chr	Cyto-band	SNP	Pos	LOD	SNPs from GWAS			
						SNP	Beta	P-value	Gene
Neuroticism									
ERF	2	2q14	rs400960	114501627	2.01				
	2	2q36	rs921280	225676062	2.32				
	3	3p14	rs1490265	67534733	**4.67**	rs332389	0.374	0.000094	SLC25A26
	7	7q36.3	rs1657290	154396352	2.26				
	11	11p15.2	rs730414	13415430	2.39				
QIMR_adolescents	10	10p14	rs2439903	6831860	2.38				
	12	12p11.2	rs2043623	29519283	2.25				
	13	13q13	rs980285	37937932	2.83				
	19	19q13.3	rs628604	59223436	**3.55**	rs7260291	0.396	0.000273	DKKL1, CCDC155
QIMR_adult	3	3p26	rs4686140	7604004	2.00				
	8	8q22.3	rs1460239	105880541	1.95				
NTR	2	2q14	D2S347	123966279	2.03				
Extraversion									
ERF	6	6q14	rs770906	83140060	3.13				
	12	12p13	rs2518144	8650712	1.98				
	17	17q25.3	rs1696754	75560132	2.74				
QIMR_adolescents	20	20p12	rs906935	5710065	2.24				
QIMR_adult	19	19q13.1	rs993983	44843320	2.08				
NTR	5	5q11.1	D5S817	11638487	3.21				
	14	14q32	ATGG002	98290290	**3.3**	rs1257657	-0.29	0.00066	c14orf177, BCL11B

Table 2 (Continued)

Study/trait	Chr	Cyto-band	SNP	Pos	LOD	SNPs from GWAS			
						SNP	Beta	P-value	Gene
ERF	7	7p14	rs43024	35267253	3.18				
	10	10p15	rs3814595	3191679	1.92				
	11	11q25	rs1824832	134313007	2.05				
	12	12q23	rs746035	103494378	2.85				
	15	15q23	rs2439378	64733880	2.16				
QIMR_adolescents	8	8q21	rs729336	85457677	2.17				
	11	11q24	rs930983	122339624	2.55				
	15	5q13	rs952121	25875398	2.96				
QIMR_adult	4	4q25	rs2046895	112466681	2.44				
NTR	12	12q24	D12S324	125191038	1.96				
Agreeableness									
ERF	4	4q32	rs2054210	164139469	2.2				
	12	12q24	rs6486532	129261499	2.7				
QIMR_adolescents	3	3p25	rs709160	12501402	**3.67**	rs2596846	-0.22	0.00059	SLC6A110
	15	15q13	rs970408	24540078	**4.07**	rs8039422	0.422	0.000377	KIAA0574
	15	15q14	rs1055356	32935394	**3.52**	rs1018325	0.242	0.00106	
Conscientiousness									
ERF	2	2q37.1	rs1868455	232103905	2.78				
	2	2q22.1	rs1402810	139217540	3.15				
	6	6p22	rs760848	18113568	2.56				
	22	22q13	rs139316	37824263	2.02				
QIMR_adolescents	7	7p22	rs1127460	652008	2.66				
	8	8q24.2	rs760327	134572411	2.63				
	9	9p23	rs1322304	10513383	1.94				
	19	19p13	rs271828	18203477	1.92				
QIMR_adult	9	9p21	rs702223	23803840	2.28				
NTR	7	7q31	D7S486	115681910	2.89				

RESULTS

Individual Scans

The results from individual genome-wide linkage scans are provided in **Supplementary Figures 1-5** and significant and suggestive findings are summarized in table 2. Significant evidence of linkage was observed for neuroticism in the ERF study at chromosome 3p14 (rs1490265, LOD = 4.67) and at chromosome 19q13 (rs628604, LOD = 3.55) in the QIMR (adolescent) sample; for extraversion at 14q32 (ATGG002, LOD = 3.3) in the NTR sample; for agreeableness at 3p25 (rs709160, LOD = 3.67) and two adjacent regions on chromosome 15 including 15q13 (rs970408, LOD = 4.07) and 15q14 (rs1055356, LOD = 3.52) in the QIMR (adolescents) study. Of these regions, we found borderline significant evidence for association after adjusting for the number of SNPs in the region for 3p14 (rs332389, p-value = $9.4*10^{-05}$, Bonferroni threshold = $2.4*10^{-05}$) and 19q13 (rs7260291, p-value = $2.7*10^{-04}$, Bonferroni threshold = $1.43*10^{-04}$) in the GWAS. Additionally for rs332389 the direction of the effect was also consistent across all the populations included in the association study[27]. Rs332389 lies in the gene *SLC25A26*, while rs7260291 is an intergenic SNP where the nearby genes include *DKKL1* and *CCDC155*. Considering the suggestive findings there was an overlap at chromosome 2q14 between ERF (LOD = 2.1) and NTR (LOD = 2.03) for neuroticism. For openness there were two adjacent regions at chromosome 11q25 for ERF (LOD = 2.05) and 11q24 for QIMR-adolescent (LOD = 2.55); and chromosome 12q23 for ERF (LOD = 2.85) and 12q24 for NTR (LOD = 1.96) that showed suggestive evidence of linkage.

Meta-analysis

Genome-wide results of the meta-analysis are illustrated in **Figure 1**. **Table 3** provides a summary of the bins with significant and suggestive evidence of linkage. None of the bins crossed the Bonferroni corrected genome-wide significance threshold. However, there were multiple bins for which both P_{SR} and P_{OR} were nominally significant (**Table 3**, **Figure 1**). For neuroticism suggestive evidence of linkage was observed for bins 18.3 and 3. The SNPs with the lowest p-values in the GWAS in these two bins were rs9875468 (bin = 3.1, p-value = $2.2*10^{-04}$) and rs12956293 (bin =18.3, p-value = $9.3*10^{-05}$) (**Figure 1A**). These associations were not significant after adjusting for the number of SNPs in the bin (**Table 3**).

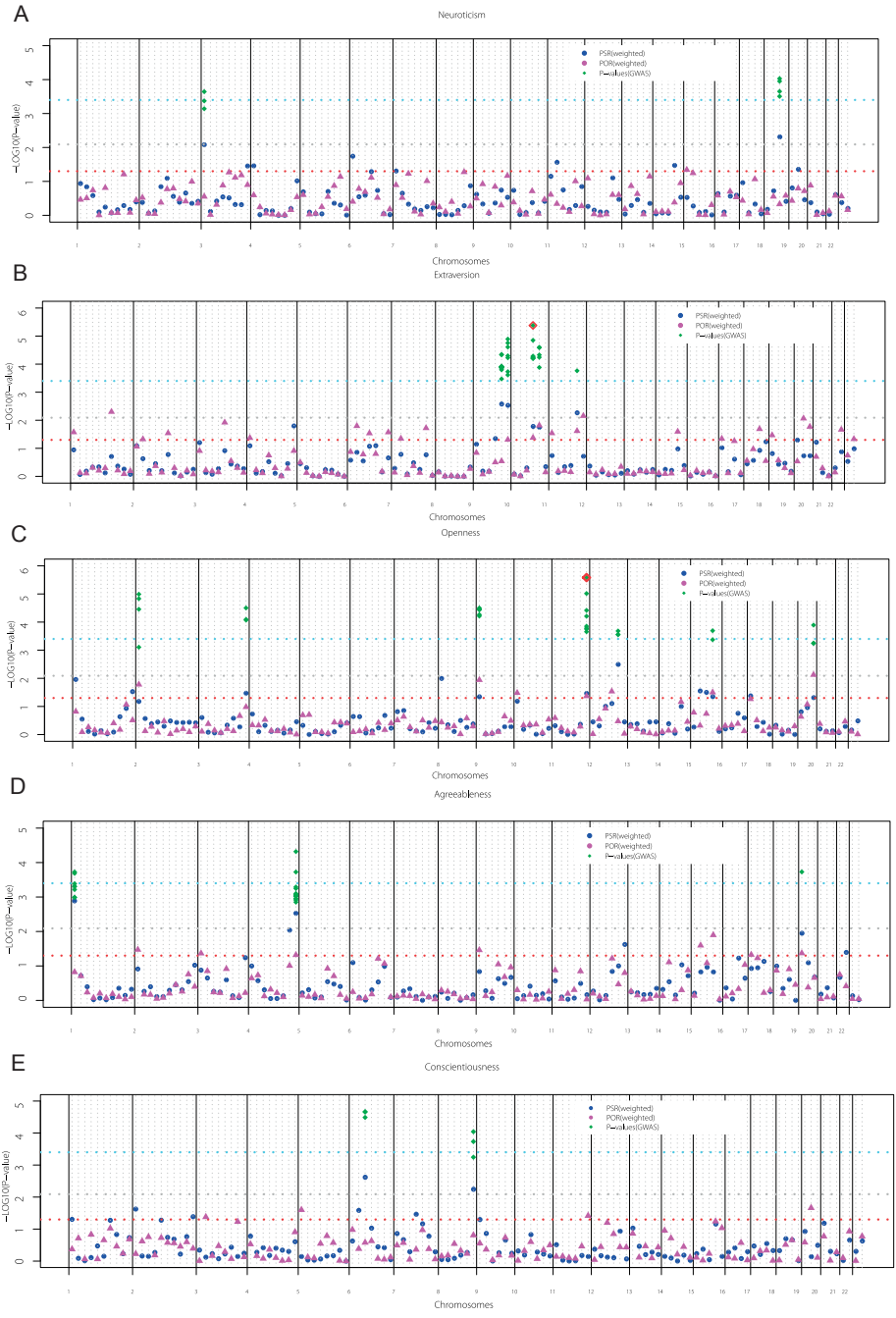

A
Neuroticism
B
Extraversion
C
Openness
D
Agreeableness
E
Conscientiousness
PSR(weighted)
POR(weighted)
P−values(GWAS)
−LOG10(P−value)
Chromosomes

Figure 1. Results of the meta-analysis of the linkage scans. Panel A for neuroticism, B for extraversion, C for openness, D for agreeableness and E for conscientiousness. Dots represent P_{SR} and triangles represent P_{OR}. The X-axis shows the whole autosomal genome divided by solid vertical lines into chromosomes, which are further divided into bins by dotted grey lines. The Y-axis shows the negative principal log of the p-values. The red horizontal dotted line represents the nominal threshold p-value = 0.05), the grey dotted horizontal line represents the suggestive threshold p-value = 0.008), and the light blue horizontal dotted line represents the bonferroni corrected threshold p-value = 0.0004). A bin is considered significant if a dot for that bin surpasses the sky blue line or if both the dot and the triangle for a specific bin are above the dotted red line. The green diamonds represent the p-values of the SNPs from the results of the meta-analysis of the GWAS of the same traits falling in the bins of interest. The green diamonds with the red outline are the p-values significant/borderline significant after correcting for the number of total SNPs in the bin.

For extraversion five bins showed nominally significant P_{SR} and P_{OR}. These included bins 9.6, 11.5 and two adjacent bins on chromosome 10 (10.4 and 10.5). Suggestive evidence of linkage was obtained for bin 9.5. Heterogeneity among studies was detected at a nominally significant level for bins 9.5, 9.6 and 10.5 (**Table 3**). The linkage signals on the adjacent bins on chromosome 9 and 10 were being caused by the same peaks that extended over 40cM (Supplementary figure 2). Adjacent bins on chromosome 9 showed evidence of suggestive linkage when the bin-width was increased to 40mbp. However, the finding for the new bin on chromosome 10 covering the previous bin 10.4 and partially covering the 10.5 bin remained significant and the adjacent bin did not show linkage signals. When comparing the significant and suggestive bins with the meta-analysis of GWAS we identified clusters of SNPs with low p-values for the bin in chromosome 10 (rs7088779, p-value = $4.2*10^{-06}$) (**Table 3**, **Figure 1B**). rs7088779 was marginally significant after being corrected for for the number of SNPs in the bin. This SNP is located between *CRTAC1* (cartilage acidic protein 1) and *C10orf28*; a region previously implicated in Alzheimer's disease.

For openness five bins were significant in that they showed both P_{SR} and $P_{OR} < 0.05$ (**Table 3**, **Figure 1C**). The bins include 9.1, 11.6, 15.4 and 19.3. Suggestive evidence of linkage was obtained for bin 12.5. A cluster of 8 SNPs with very low p-vales was identified from the GWAS for the bin 11.6 (**Figure 1C**), which maps to 11q24 region. The most significant SNP (rs677035) showed an observed p-value of $2.6*10^{-06}$ which passed the Bonferroni threshold. Rs677035 is an intergenic SNP located between *FLI1* and *KCNJ1*. *KCNJ1* is potassium inwardly-rectifying channel, subfamily J, member 1 and *FLI1* encodes the Friend leukemia virus integration 1 protein. No heterogeneity was detected for any of the bins that showed nominally significant P_{SR} and P_{OR}.

Two bins showed nominally significant P_{SR} and P_{OR} for agreeableness. These include bins on chromosome 4 and 19 (**Table 3**, **Figure 1D**). Suggestive evidence of linkage was obtained for bin 1.1. Heterogeneity between studies was detected for bin 4.8. We identified clusters of SNPs with low p-values from the GWAS for bins 1.1 (rs6699411, p-value = $1.9*10^{-04}$) and 4.8 (rs13113475, p-value = $4.8*10^{-05}$) and an isolated SNP (rs10413538) with a low p-value ($3.7*10^{-04}$) for bin 19.1 (**Figure 1D**). These SNPs were not significant when corrected for the total number of SNPs in the bin.

Suggestive linkage was observed for bins 6.3 and 8.6 for conscientiousness (**Table 3**, **Figure 1E**). Bin 6.3 showed strong evidence of heterogeneity among studies. We identified a cluster of 3 SNPs for bin 6.3 in the GWAS with the most significant (rs2807510) having a p-value of $2.18*10^{-05}$, and a cluster of 3 SNPs for bin 8.6 with the most significant SNP (rs2977475) with a p-value of $8.33*10^{-05}$. None of these SNPs surpassed Bonferroni corrected threshold.

Table 3: Comparison of interesting regions of Linkage meta-analysis with the Personality GWAS meta-analysis

Trait	Bin†	cyto	Summed Rank	SR pvalue	OR pvalue	P_{HET}	Approx. region (Mbp)	SNP in GWAS	β	P-value (nominal)	Bonferroni Threshold[a]	gene	left_gene	right_gene
Neuroticism	3.1	3p25-26	419.5	0.008	0.280	0.045	0-25	rs9875468	-0.69	$2.2*1^{-04}$	$1.7*1^{-06}$	SRGAP3	RAD18	LOC391508
	18.3	18q22	429.6	0.005	0.468	0.021	50-75	rs12956293	0.67	$9.3*1^{-05}$	$1.7*1^{-06}$	intergenic	LOC100132647	CBLN2
Extraversion	9.5	9q31-32	439.8	0.003	0.287	0.043	100-125	rs10817056	0.36	$4.5*1^{-05}$	$1.7*1^{-06}$	intergenic	SVEP1	MUSK
	9.6	9q34	438.2	0.003	0.049	0.036	125-150	rs11103814	-0.34	$1.3*1^{-05}$	$4.2*1^{-06}$	intergenic	LOC401557	C9orf62
	10.4	10q24	403.6	0.017	0.042	0.063	75-100	rs7088779	-0.48	*$4.2*1^{-06}$*	$2.1*1^{-06}$	intergenic	CRTAC1	C10orf28
	10.5	10q24	401.9	0.018	0.015	0.042	100-125	rs7905091	0.5	$2.5*1^{-05}$	$2.1*1^{-06}$	intergenic	LOXL4	C10orf33
	11.5	11q22	427.6	0.005	0.024	0.073	100-125	rs12419614	1.73	$1.1*1^{-04}$	$2.0*1^{-06}$	GRIA4	OR2AL1P	KIAA1826
Openness	2.1	2p25	362.3	0.067	0.016	0.186	0-25	rs11685318	-0.38	$1.0*1^{-05}$	$1.8*1^{-06}$	intergenic	PDIA6	KCNF1
	9.1	9p21	375.3	0.045	0.011	0.053	0-25	rs10966354	0.61	$2.6*1^{-05}$	$1.4*1^{-06}$	intergenic	LOC402360	C9orf82
	11.6	11q24	383.7	0.034	0.043	0.625	125-150	rs677035	-0.56	***$2.6*1^{-06}$***	$4.4*1^{-06}$	intergenic	FLI1	KCNJ1
	12.5	12q24	436.8	0.003	0.335	0.155	100-125	rs12297453	-0.31	$2.0*1^{-04}$	$2.3*1^{-06}$	CCDC60	LOC144742	LOC100132232
	15.4	15q26	375.7	0.045	0.031	0.280	75-100	rs7175820	0.48	$2.7*1^{-04}$	$2.0*1^{-06}$	intergenic	LOC390638	SV2B
	19.3	19q13	373.2	0.048	0.007	0.165	50-75	rs2287826	-0.6	$3.4*1^{-04}$	$4.8*1^{-06}$	LAIR1	LILRA4	LAIR1
Agreeableness	1.1	1p36	449.7	0.001	0.151	0.016	0-25	rs6699411	-0.28	$1.9*1^{-04}$	$2.6*1^{-06}$	RP1	PRDM2	TBCAP2
	4.8	4q34	437.9	0.003	0.048	0.008	175-200	rs13113475	0.28	$4.8*1^{-05}$	$3.1*1^{-06}$	intergenic	SPCS3	VEGFC
	19.1	19p13	412.5	0.011	0.042	0.134	0-25	rs10413538	0.55	$3.7*1^{-04}$	$3.6*1^{-06}$	ZNF826	LOC100131306	ZNF737
Conscientiousness	6..3	6q13	440.9	0.002	0.263	0.002	50-75	rs2807510	-1.49	$2.1*1^{-05}$	$2.2*1^{-06}$	RIMS1	KRT19P1	LOC643067
	8.6	8q24	426.4	0.006	0.155	0.088	125-150	rs2977475	-0.33	$8.3*1^{-05}$	$2.2*1^{-06}$	EIF2C2	CHRAC1	LOC100130939

[a] bonferroni threshold based on the number of SNPs from the GWAS present in the bin p-values from GWAS significant after bonferroni correction in bold italics. Border line significant in italics only.

DISCUSSION

In the individual studies we found significant evidence for linkage for neuroticism to chromosome 3p14 (rs1490265, LOD = 4.67) and to chromosome 19q13 (rs628604, LOD = 3.55), for extraversion to 14q32 (ATGG002, LOD = 3.3); for agreeableness to 3p25 (rs709160, LOD = 3.67) and two adjacent regions on chromosome 15 including 15q13 (rs970408, LOD = 4.07) and 15q14 (rs1055356, LOD = 3.52). In the meta-analysis, we found strong evidence of linkage of extraversion to 9q34, 10q24 and 11q22, openness to 2p25, 3q26, 9p21, 11q24, 15q26 and 19q13, and agreeableness to 4q34 and 19p13. When combining these data with the association findings of the recently published GWAS of these personality traits, there was significant evidence of association between openness and rs677035 (p-value = $2.6*10^{-06}$) in the 11q24 region, which retained significance after adjusting for multiple testing. The evidence was borderline for association between neuroticism and rs332389 (p-value = $9*10^{-05}$) in the 3p14 region and between extraversion and rs7088779 (p-value = $4.2*10^{-06}$) in the 10q24 region.

In this study we performed genome-wide linkage analysis of the NEO personality traits in four independent populations and then combined these together in a meta-analysis using the GSMA method. Our meta-analysis included 6149 individuals for multiple extended families and families with sib-ships. There are a number of methodological issues relevant for the interpretation of the findings. For the meta-analysis we used the physical map (base pair positions) as opposed to the genetic map (centiMorgan position) that was used in previous studies that applied GSMA to meta-analyze linkage scans. The physical map is more accurate and therefore better suited to defining the bin boundaries. We also used a bin-width of 25 million base pairs (which would roughly translate to 25 cM) as opposed to the traditional 30cM bin-width used in all previous studies. Our choice of 25 mbp was made after taking into account the genetic maps of all four studies and specially the position of last reported markers on each chromosome in all studies (on chromosome 22 the last reported marker was located at around 49mbp). This helped us avoid any sort of manipulation of the results from any population and use the maximum available information. However, this selection could lead to a correlation between two adjacent bins. Interestingly this correlation was observed only for extraversion in two pairs of adjacent bins on chromosomes 9 and 10. The size of the linkage peak on chromosome 9 extends to about 50 cM (supplementary figure 2) implying that choice of a bin-width of more than 30 cM could not have removed this correlation. For chromosome 10 the significance remained even after increasing the bin-width to 40mbp, which suggests that our result on 10q24 for extraversion is consistent. The GSMA gave broad linkage regions spread over 25 mbp, but we made an effort to localize the susceptibility genes by using additional information from the meta-analysis of GWAS.

It is interesting to note that none of the significant regions from the individual scans showed any evidence of linkage in the meta-analysis. There may be several explanations for this

inconsistency. First, our significant findings of individual studies may just be false positives. This may be a possibility, but it is difficult to believe that all significant findings are false, since variance components linkage analysis is usually robust and conservative in detecting linkage[57] given the trait is normally distributed (which was true in our case). Second, there may have been differences in the power of the various studies. For example, the reason why the linkage of neuroticism to 3p14 was detected in the ERF study only may be the size of the study, which was twice as large as any other included study. Third, there may be locus heterogeneity across populations. High locus heterogeneity, which results in inconsistent linkage peaks, is one characteristic of complex traits like personality. It has been recognized by the authors of the GSMA method that it cannot deal with heterogeneity across studies and as such population-specific/rare mutations, which are more likely to have larger effects, will not be detected in the meta-analysis[53] with this method. When considering our results, the rank based test used is insensitive to the significance of a linkage peak in an individual study and is more adept at finding subtle linkage peaks present in all studies included in the meta-analysis. For instance the meta-analysis ignores the overlap at chromosome 2q14 between ERF (LOD = 2.01) and NTR (LOD = 2.03) for neuroticism despite falling in the same bin. Similarly two adjacent regions for openness at chromosome 11q25 for ERF (LOD = 2.05) and 11q24 for QIMR-adolescent (LOD = 2.55) and chromosome 12q23 for ERF (LOD = 2.85) and 12q24 for NTR (LOD = 1.96) that showed suggestive evidence of linkage were not picked up in the meta-analysis. Such results were also observed in studies that used parametric methods (Fisher's method) to meta-analyze genome-wide linkage scans[58].

Our major finding from the individual scans was 3p14 for neuroticism, which gave very significant linkage signals (LOD = 4.67) in the ERF study, showed strong evidence of association (rs332389, p-value = $9.4*10^{-05}$) in the GWAS and was previously linked to neuroticism in a linkage scan of nicotine dependent smokers[59]. Rs332389 is an intronic (NM_173471.2) SNP that lies in the gene *SLC25A26*. *SLC25A26* encodes S-adenosylmethionine mitochondrial carrier protein[60-61]. S-adenosylmethionine is a substrate involved in methyl group transfers and is orally administered to treat depression[62-63]. The most interesting finding of the meta-analysis is 11q24 that showed nominally significant P_{SR} and P_{OR} and also significant association in the GWAS after adjusting for multiple testing. The 11q24 region is implicated in mental retardation[64] and migraine[65]. The significant SNP in the GWAS rs677035 is located between *FLI1* and *KCNJ1*. Previously we have found linkage to other potassium channel genes including *KCNJ2*, *KCNJ6* and *KCNJ16* (see Chapter 3.1), making *KCNJ1* the most interesting gene in the region. The other regions of interest include 10q24 for extraversion. The region showed both P_{SR} and $P_{OR} < 0.05$, and additionally showed borderline significant evidence of association in the meta-analysis of GWAS after adjusting for multiple testing. Both regions have been implicated in neurological diseases earlier. At present we cannot discard other regions that showed a high probability of containing true susceptibility loci in the linkage analyses as being false positives based on insignificant

association results of the GWAS. Our GWAS data have been relatively underpowered. These regions are good candidates for sequencing exomes and regulatory regions and may unveil variants with moderate to large effects contributing to the make-up of human personality. All eyes are now on the next generation sequencing techniques, which are expected to uncover these rare variants with relatively high effects on trait outcomes. These developments have brought family based studies back to the frontline of research. However, our study forecasts a number of the problems to be anticipated in high-throughput sequencing in regions of interest. In particular, replications may prove difficult across populations. At present, the methods combining evidence from different family based studies have major limitations. The criteria for significant evidence for linkage in the meta-analyses were not reached in this paper combining large population-based studies. The highest level of significance for linkage reached, translated into a high probability (of unknown size) to include a susceptibility locus. It is obvious that false negatives may occur due to locus heterogeneity using the current methods. To overcome these problems, we combined the data from the linkage analyses of the five basic NEO personality traits in four independent populations with those of the genome wide association analyses. Most likely, also in the analyses of next generation sequencing this combination of linkage and association will be crucial. Here we relied on classical association analysis and genetic imputations based on HapMAP. This approach will be strengthened by imputing populations that underwent GWAS using the 1000 genomes, providing a greater resolution and a better coverage of rare variants.

In summary, we found significant evidence of linkage of neuroticism to 3p14 and regions with high probability of containing true susceptibility loci including 11q24 for openness and 10q24 for extraversion. An important conclusion from our study is, however, that the current methods available for meta-analyzing family based studies are far from optimal. However, we also conclude that combining information from linkage and association is a useful approach in the identification of genes involved in complex phenotypes such as personality.

Acknowledgements

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 Organisation 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.

NTR: We acknowledge financial support from the Netherlands Organization for Scientific Research (NWO): grants 575-25-006, 480-04-004, 904-61-090; 904-61-193, 400-05-717 and Spinozapremie SPI 56-464-14192. M.H.M.deM. is financially supported by ZonMW (Addiction) Grant number 311-60008. We further acknowledge financial support from the Center for Medical Systems Biology (NWO Genomics), Neuroscience Campus Amsterdam; EU/QLRT-2001-01254; NIMH R01 MH059160; Geestkracht program of ZonMW (10-000-1002); matching funds from universities and mental health care institutes involved in NESDA. Genotyping was funded by the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health, and analysis was supported by grants from GAIN and the NIMH (MH081802). Genotype data were obtained from dbGaP (http://www.ncbi.nlm.nih.gov/dbgap, accession number phs000020.v1.p1).

QIMR: We thank Marlene Grace and Ann Eldridge for sample collection; Anjali Henders, Megan Campbell, Lisa Bowdler, Steven Crooks and staff of the Molecular Epidemiology Laboratory for sample processing and preparation; Harry Beeby, David Smyth and Daniel Park for IT support. We acknowledge support from the Australian Research Council (A7960034, A79906588, A79801419, DP0212016, DP0343921), Beyond Blue, and the Borderline Personality Disorder Research Foundation. Genotyping was funded by the National Health and Medical Research Council (Medical Bioinformatics Genomics Proteomics Program, 389891). Further, we gratefully acknowledge Drs Dale R Nyholt and especially Scott Gordon for their substantial efforts involving the QC and preparation of the QIMR and NAG/IRPG GWA datasets from which SNPs were selected. Dr Nyholt also contributed 8% of the NAG/IRPG GWA cohort (NHMRC IDs 339462, 442981, 389938, 496739). In addition, we thank him for helpful discussions regarding methodology.

REFERENCES

1. Bienvenu, O.J., Hettema, J.M., Neale, M.C., Prescott, C.A., and Kendler, K.S. (2007). Low extraversion and high neuroticism as indices of genetic and environmental risk for social phobia, agoraphobia, and animal phobia. Am J Psychiatry 164, 1714-1721.
2. Cloninger, C.R., Svrakic, D.M., and Przybeck, T.R. (1993). A psychobiological model of temperament and character. Arch Gen Psychiatry 50, 975-990.
3. Cramer, D. (1991). Type A behaviour pattern, extraversion, neuroticism and psychological distress. Br J Med Psychol 64 (Pt 1), 73-83.
4. Crowe, M., Andel, R., Pedersen, N.L., Fratiglioni, L., and Gatz, M. (2006). Personality and risk of cognitive impairment 25 years later. Psychol Aging 21, 573-580.
5. Fanous, A.H., Neale, M.C., Aggen, S.H., and Kendler, K.S. (2007). A longitudinal study of personality and major depression in a population-based sample of male twins. Psychol Med 37, 1163-1172.
6. Goodwin, R.D., Castro, M., and Kovacs, M. (2006). Major depression and allergy: does neuroticism explain the relationship? Psychosom Med 68, 94-98.
7. Hettema, J.M., Neale, M.C., Myers, J.M., Prescott, C.A., and Kendler, K.S. (2006). A population-based twin study of the relationship between neuroticism and internalizing disorders. Am J Psychiatry 163, 857-864.
8. Kaasinen, V., Nurmi, E., Bergman, J., Eskola, O., Solin, O., Sonninen, P., and Rinne, J.O. (2001). Personality traits and brain dopaminergic function in Parkinson's disease. Proc Natl Acad Sci U S A 98, 13272-13277.
9. McCrae, R.R., Yang, J., Costa, P.T., Jr., Dai, X., Yao, S., Cai, T., and Gao, B. (2001). Personality profiles and the prediction of categorical personality disorders. J Pers 69, 155-174.
10. Menza, M. (2000). The personality associated with Parkinson's disease. Curr Psychiatry Rep 2, 421-426.
11. Sandahl, C., Lindberg, S., and Bergman, H. (1987). The relation between drinking habits and neuroticism and weak ego among male and female alcoholic patients. Acta Psychiatr Scand 75, 500-508.
12. Van Os, J., and Jones, P.B. (2001). Neuroticism as a risk factor for schizophrenia. Psychol Med 31, 1129-1134.
13. Wilson, R.S., Schneider, J.A., Arnold, S.E., Bienias, J.L., and Bennett, D.A. (2007). Conscientiousness and the incidence of Alzheimer disease and mild cognitive impairment. Arch Gen Psychiatry 64, 1204-1212.
14. Rahman, M.A., and Eysenck, S.B. (1978). Psychoticism and response to treatment in neurotic patients. Behav Res Ther 16, 183-189.
15. Meins, W., and Dammast, J. (2000). Do personality traits predict the occurrence of Alzheimer's disease? Int J Geriatr Psychiatry 15, 120-124.
16. Distel, M.A., Trull, T.J., Willemsen, G., Vink, J.M., Derom, C.A., Lynskey, M., Martin, N.G., and Boomsma, D.I. (2009). The five-factor model of personality and borderline personality disorder: a genetic analysis of comorbidity. Biol Psychiatry 66, 1131-1138.
17. Goodwin, R.D., and Stein, M.B. (2003). Peptic ulcer disease and neuroticism in the United States adult population. Psychother Psychosom 72, 10-15.
18. Lichtenstein, P., Pedersen, N.L., Plomin, R., Faire, U.D., and McClearn, G.E. (1989). Type A behavior pattern, related personality traits and self-reported coronary heart disease. Person individ Diff 10, 419-426.
19. Robertson, D.A., Ray, J., Diamond, I., and Edwards, J.G. (1989). Personality profile and affective state of patients with inflammatory bowel disease. Gut 30, 623-626.
20. Shipley, B.A., Weiss, A., Der, G., Taylor, M.D., and Deary, I.J. (2007). Neuroticism, extraversion, and mortality in the UK Health and Lifestyle Survey: a 21-year prospective cohort study. Psychosom Med 69, 923-931.
21. Winstow, D.J., Wakefield, J.A., and Goldsmith, W.M. (1990). The relationship between personality, health symptoms and disease. Person individ Diff 11, 717-723.
22. Sutin, A.R., Costa, P.T., Jr., Uda, M., Ferrucci, L., Schlessinger, D., and Terracciano, A. (2010). Personality and metabolic syndrome. Age (Dordr) 32, 513-519.
23. Terracciano, A., Sutin, A.R., McCrae, R.R., Deiana, B., Ferrucci, L., Schlessinger, D., Uda, M., and Costa, P.T., Jr. (2009). Facets of personality linked to underweight and overweight. Psychosom Med 71, 682-689.

24. Sutin, A.R., Scuteri, A., Lakatta, E.G., Tarasov, K.V., Ferrucci, L., Costa, P.T., Jr., Schlessinger, D., Uda, M., and Terracciano, A. (2010). Trait antagonism and the progression of arterial thickening: women with antagonistic traits have similar carotid arterial thickness as men. Hypertension 56, 617-622.

25. Sutin, A.R., Terracciano, A., Deiana, B., Uda, M., Schlessinger, D., Lakatta, E.G., and Costa, P.T., Jr. (2010). Cholesterol, triglycerides, and the Five-Factor Model of personality. Biol Psychol 84, 186-191.

26. Sutin, A.R., Terracciano, A., Deiana, B., Naitza, S., Ferrucci, L., Uda, M., Schlessinger, D., and Costa, P.T., Jr. (2010). High neuroticism and low conscientiousness are associated with interleukin-6. Psychol Med 40, 1485-1493.

27. de Moor, M.H.M., Costa, P.T., Terracciano, A., Krueger, R.F., Geus, E.d., Toshiko, T., Penninx, B.W.J.H., Esko, T., Madden, P.A.F., Derringer, J., et al. (2010). Meta-analysis of genome-wide association studies for personality. Mol Psychiatry.

28. Maher, B. (2008). Personal genomes: The case of the missing heritability. Nature 456, 18-21.

29. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42, 565-569.

30. Dickson, S.P., Wang, K., Krantz, I., Hakonarson, H., and Goldstein, D.B. (2010). Rare variants create synthetic genome-wide associations. PLoS Biol 8, e1000294.

31. Sham, P.C., Cherny, S.S., Purcell, S., and Hewitt, J.K. (2000). Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. Am J Hum Genet 66, 1616-1630.

32. Costa, P.T., and McCrae, R.R. (1992). Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI) Professional Manual. Psychological Assessment Resources.

33. McCrae, R.R., Terracciano, A., and Personality Profiles of Cultures, P. (2005). Universal features of personality traits from the observer's perspective: data from 50 cultures. J Pers Soc Psychol 88, 547-561.

34. Terracciano, A., Costa, P.T., Jr., and McCrae, R.R. (2006). Personality plasticity after age 30. Pers Soc Psychol Bull 32, 999-1009.

35. Biesanz, J.C., and West, S.G. (2004). Towards understanding assessments of the big five: multitrait-multimethod analyses of convergent and discriminant validity across measurement occasion and type of observer. J Pers 72, 845-876.

36. Costa, P.T., Jr., Terracciano, A., and McCrae, R.R. (2001). Gender differences in personality traits across cultures: robust and surprising findings. J Pers Soc Psychol 81, 322-331.

37. Budaev, S.V. (1999). Personality and Individual Differences. Elsevier 26, 801-813.

38. Goodwin, R.D., and Gotlib, I.H. (2004). Gender differences in depression: the role of personality factors. Psychiatry Res 126, 135-142.

39. Pardo, L.M., MacKay, I., Oostra, B., van Duijn, C.M., and Aulchenko, Y.S. (2005). The effect of genetic drift in a young genetically isolated population. Ann Hum Genet 69, 288-295.

40. Miller, S.A., Dykes, D.D., and Polesky, H.F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16, 1215.

41. Abecasis, G.R., Cherny, S.S., Cookson, W.O., and Cardon, L.R. (2002). Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30, 97-101.

42. Falchi, M., and Fuchsberger, C. (2008). Jenti: an efficient tool for mining complex inbred genealogies. Bioinformatics 24, 724-726.

43. Axenovich, T.I., Zorkoltseva, I.V., Liu, F., Kirichenko, A.V., and Aulchenko, Y.S. (2008). Breaking loops in large complex pedigrees. Hum Hered 65, 57-65.

44. Bellenguez, C., Ober, C., and Bourgain, C. (2009). A multiple splitting approach to linkage analysis in large pedigrees identifies a linkage to asthma on chromosome 12. Genet Epidemiol 33, 207-216.

45. Boomsma, D.I., de Geus, E.J., Vink, J.M., Stubbe, J.H., Distel, M.A., Hottenga, J.J., Posthuma, D., van Beijsterveldt, T.C., Hudziak, J.J., Bartels, M., et al. (2006). Netherlands Twin Register: from twins to twin families. Twin Res Hum Genet 9, 849-857.

46. Matise, T.C., Chen, F., Chen, W., De La Vega, F.M., Hansen, M., He, C., Hyland, F.C., Kennedy, G.C., Kong, X., Murray, S.S., et al. (2007). A second-generation combined linkage physical map of the human genome. Genome Res 17, 1783-1786.

47. Pergadia, M.L., Agrawal, A., Loukola, A., Montgomery, G.W., Broms, U., Saccone, S.F., Wang, J.C., Todorov, A.A., Heikkila, K., Statham, D.J., et al. (2009). Genetic linkage findings for DSM-IV nicotine withdrawal in two populations. Am J Med Genet B Neuropsychiatr Genet 150B, 950-959.

48. Pergadia, M.L., Agrawal, A., Loukola, A., Montgomery, G.W., Broms, U., Saccone, S.F., Wang, J.C., Todorov, A.A., Keikkila, K., Statham, D., et al. (2009). Genetic linkage findings for DSM-IV nicotine withdrawal in two populations. American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 150B, 950-959.

49. Saccone, S.F., Pergadia, M.L., Broms, U., Loukola, A., Montgomery, G.W., Wang, J.C., Agrawal, A., Rice, J.P., Dick, D.M., Todd, R.D., et al. (2007). Genetic linkage to chromosome 22q12 for a heavy smoking quantitative trait in two independent samples. Am J Hum Genet 80, 856-866.

50. Wright, M.J., and Martin, N.G. (2004). The Brisbane Adolescent Twin Study: outline of study methods and research projects. Australian Journal of Psychology 56, 65-78.

51. Distel, M.A., Trull, T.J., Derom, C.A., Thiery, E.W., Grimmer, M.A., Martin, N.G., Willemsen, G., and Boomsma, D.I. (2007). Heritability of borderline personality disorder features is similar across three countries. Psychol Med 38, 1219-1229.

52. Medland, S.E., Nyholt, D.R., Painter, J.N., McEvoy, B.P., McRae, A.F., Zhu, G., Gordon, S.D., Ferreira, M.A., Wright, M.J., Henders, A.K., et al. (2009). Common variants in the trichohyalin gene are associated with straight hair in Europeans. Am J Hum Genet 85, 750-755.

53. Wise, L.H., Lanchbury, J.S., and Lewis, C.M. (1999). Meta-analysis of genome searches. Ann Hum Genet 63, 263-272.

54. Pardi, F., Levinson, D.F., and Lewis, C.M. (2005). GSMA: software implementation of the genome search meta-analysis method. Bioinformatics 21, 4430-4431.

55. Levinson, D.F., Levinson, M.D., Segurado, R., and Lewis, C.M. (2003). Genome scan meta-analysis of schizophrenia and bipolar disorder, part I: Methods and power analysis. Am J Hum Genet 73, 17-33.

56. Zintzaras, E., and Ioannidis, J.P. (2005). HEGESMA: genome search meta-analysis and heterogeneity testing. Bioinformatics 21, 3672-3673.

57. Shugart, Y.Y., O'Connell, J.R., and Wilson, A.F. (2002). An evaluation of the variance components approach: type I error, power and size of the estimated effect. Eur J Hum Genet 10, 133-136.

58. Province, M.A., Kardia, S.L., Ranade, K., Rao, D.C., Thiel, B.A., Cooper, R.S., Risch, N., Turner, S.T., Cox, D.R., Hunt, S.C., et al. (2003). A meta-analysis of genome-wide linkage scans for hypertension: the National Heart, Lung and Blood Institute Family Blood Pressure Program. Am J Hypertens 16, 144-147.

59. Neale, B.M., Sullivan, P.F., and Kendler, K.S. (2005). A genome scan of neuroticism in nicotine dependent smokers. Am J Med Genet B Neuropsychiatr Genet 132, 65-69.

60. Agrimi, G., Di Noia, M.A., Marobbio, C.M., Fiermonte, G., Lasorsa, F.M., and Palmieri, F. (2004). Identification of the human mitochondrial S-adenosylmethionine transporter: bacterial expression, reconstitution, functional characterization and tissue distribution. Biochem J 379, 183-190.

61. Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., et al. (2004). Complete sequencing and characterization of 21,243 full-length human cDNAs. Nat Genet 36, 40-45.

62. Kagan, B.L., Sultzer, D.L., Rosenlicht, N., and Gerner, R.H. (1990). Oral S-adenosylmethionine in depression: a randomized, double-blind, placebo-controlled trial. Am J Psychiatry 147, 591-595.

63. Rosenbaum, J.F., Fava, M., Falk, W.E., Pollack, M.H., Cohen, L.S., Cohen, B.M., and Zubenko, G.S. (1990). The antidepressant potential of oral S-adenosyl-l-methionine. Acta Psychiatr Scand 81, 432-436.

64. Bhalla, K., Luo, Y., Buchan, T., Beachem, M.A., Guzauskas, G.F., Ladd, S., Bratcher, S.J., Schroer, R.J., Balsamo, J., DuPont, B.R., et al. (2008). Alterations in CDH15 and KIRREL3 in patients with mild to severe intellectual disability. Am J Hum Genet 83, 703-713.

65. Cader, Z.M., Noble-Topham, S., Dyment, D.A., Cherny, S.S., Brown, J.D., Rice, G.P., and Ebers, G.C. (2003). Significant linkage to migraine with aura on chromosome 11q24. Hum Mol Genet 12, 2511-2517.

66. Storey, E., Bahlo, M., Fahey, M., Sisson, O., Lueck, C.J., and Gardner, R.J. (2009). A new dominantly inherited pure cerebellar ataxia, SCA 30. J Neurol Neurosurg Psychiatry 80, 408-411.

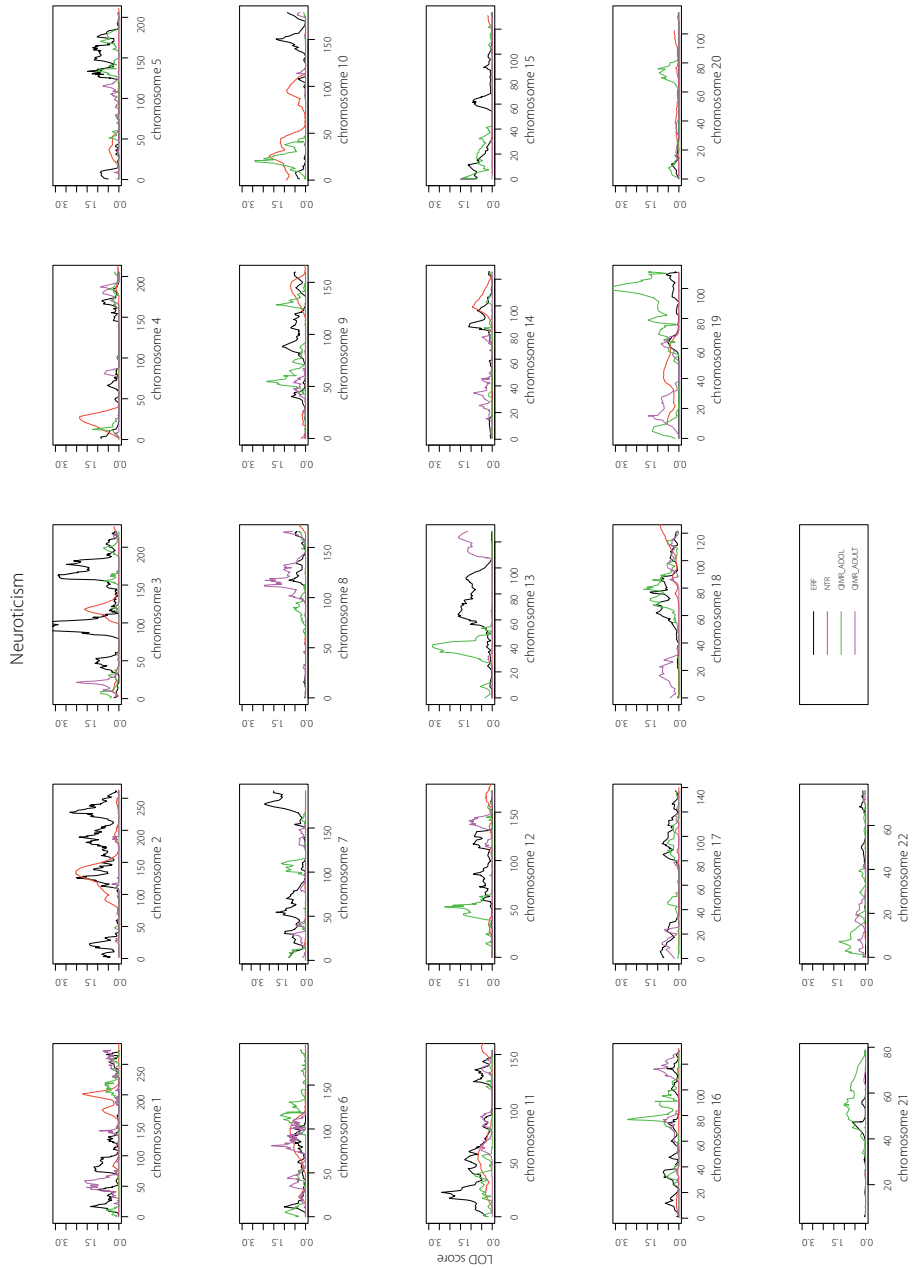
67. Nikali, K., Suomalainen, A., Terwilliger, J., Koskinen, T., Weissenbach, J., and Peltonen, L. (1995). Random search for shared chromosomal regions in four affected individuals: the assignment of a new hereditary ataxia locus. Am J Hum Genet 56, 1088-1095.

68. Liu, F., Arias-Vasquez, A., Sleegers, K., Aulchenko, Y.S., Kayser, M., Sanchez-Juan, P., Feng, B.J., Bertoli-Avella, A.M., van Swieten, J., Axenovich, T.I., et al. (2007). A genomewide screen for late-onset Alzheimer disease in a genetically isolated Dutch population. Am J Hum Genet 81, 17-31.

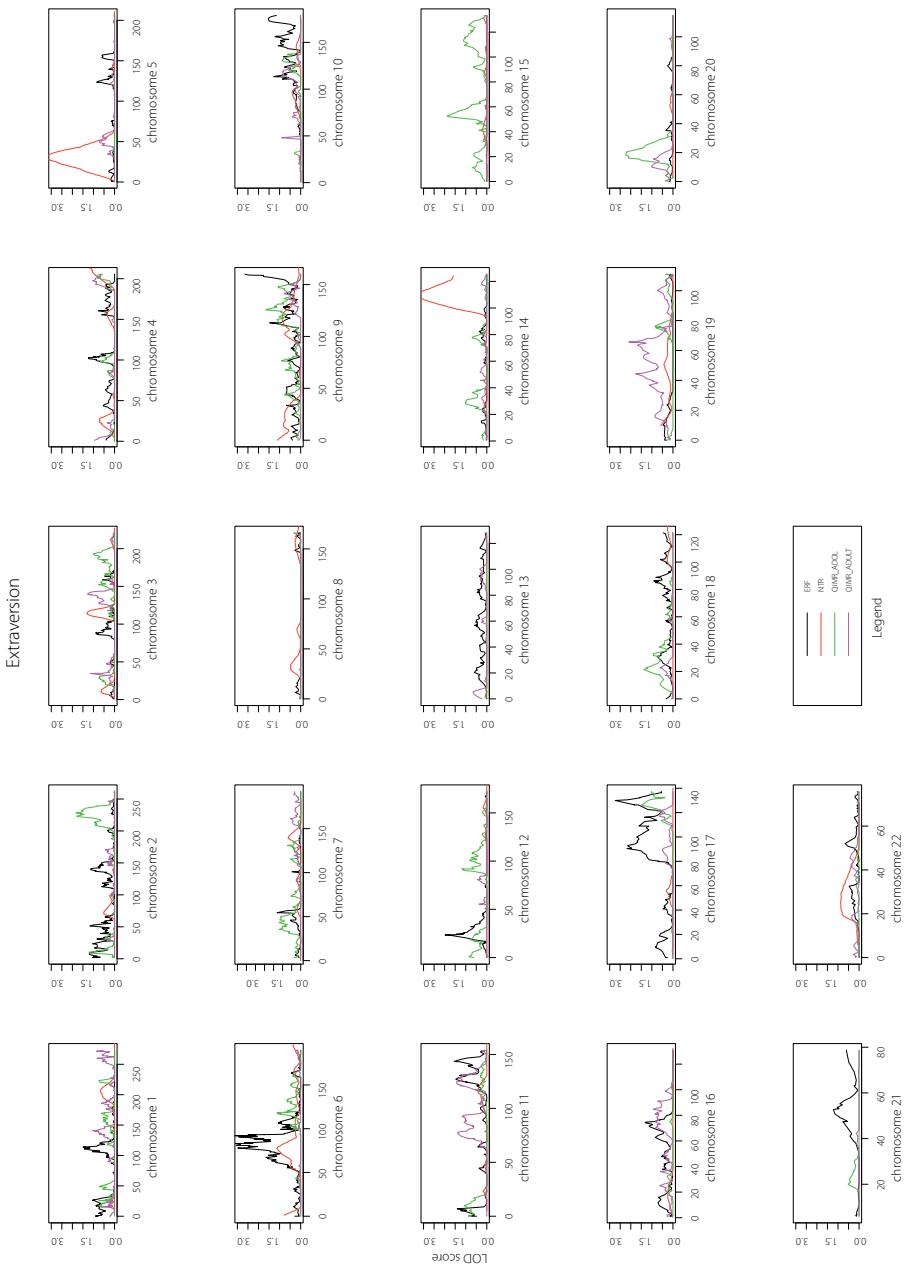
69. Bertram, L., Blacker, D., Mullin, K., Keeney, D., Jones, J., Basu, S., Yhu, S., McInnis, M.G., Go, R.C., Vekrellis, K., et al. (2000). Evidence for genetic linkage of Alzheimer's disease to chromosome 10q. Science 290, 2302-2303.

SUPPLEMENTARY INFORMATION

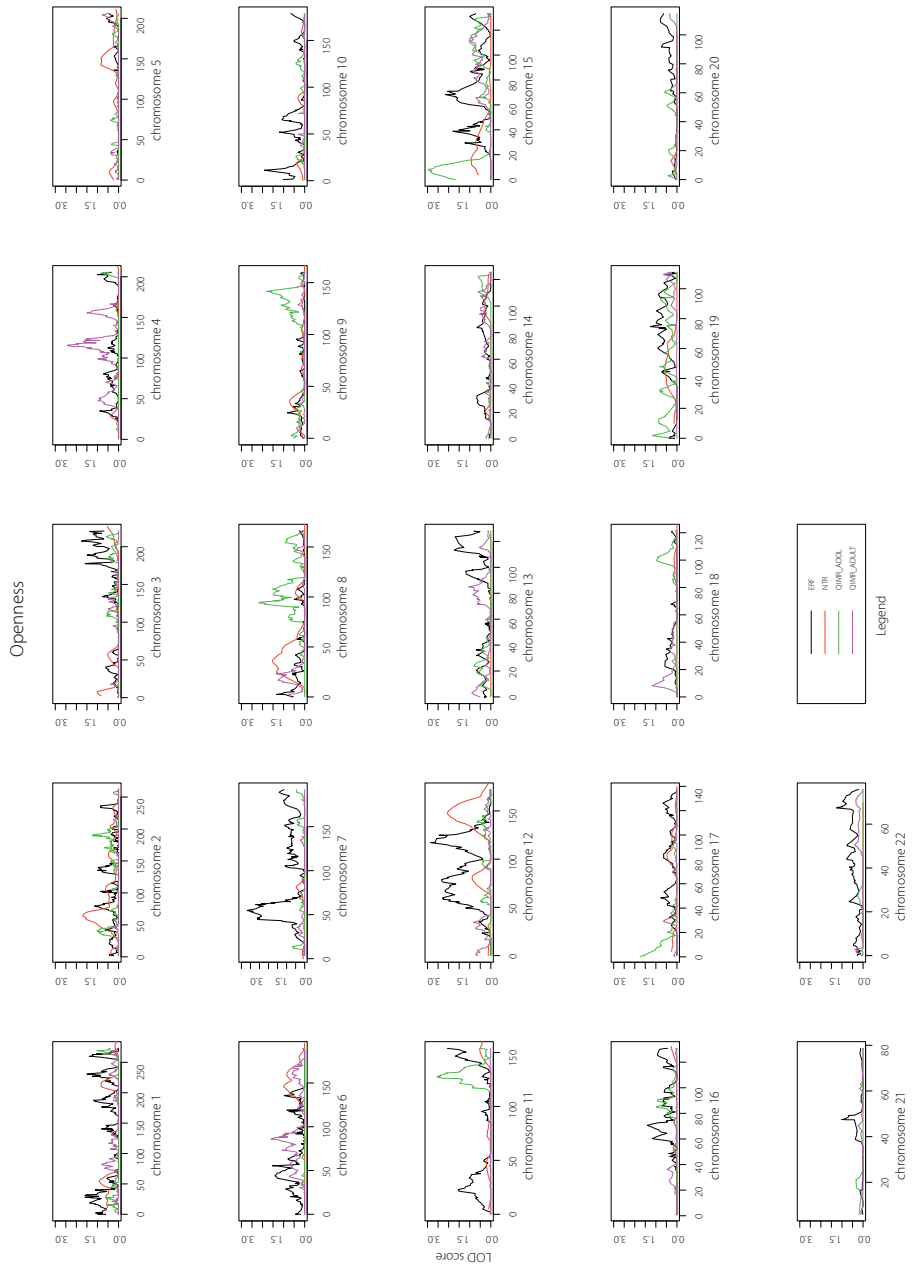
Supplementary Figure 1. Genome-wide linkage plots for individual scans for neuroticism. The vertical axis shows the log of odds score and the horizontal axis shows the distance on the chromosome in centiMorgan.



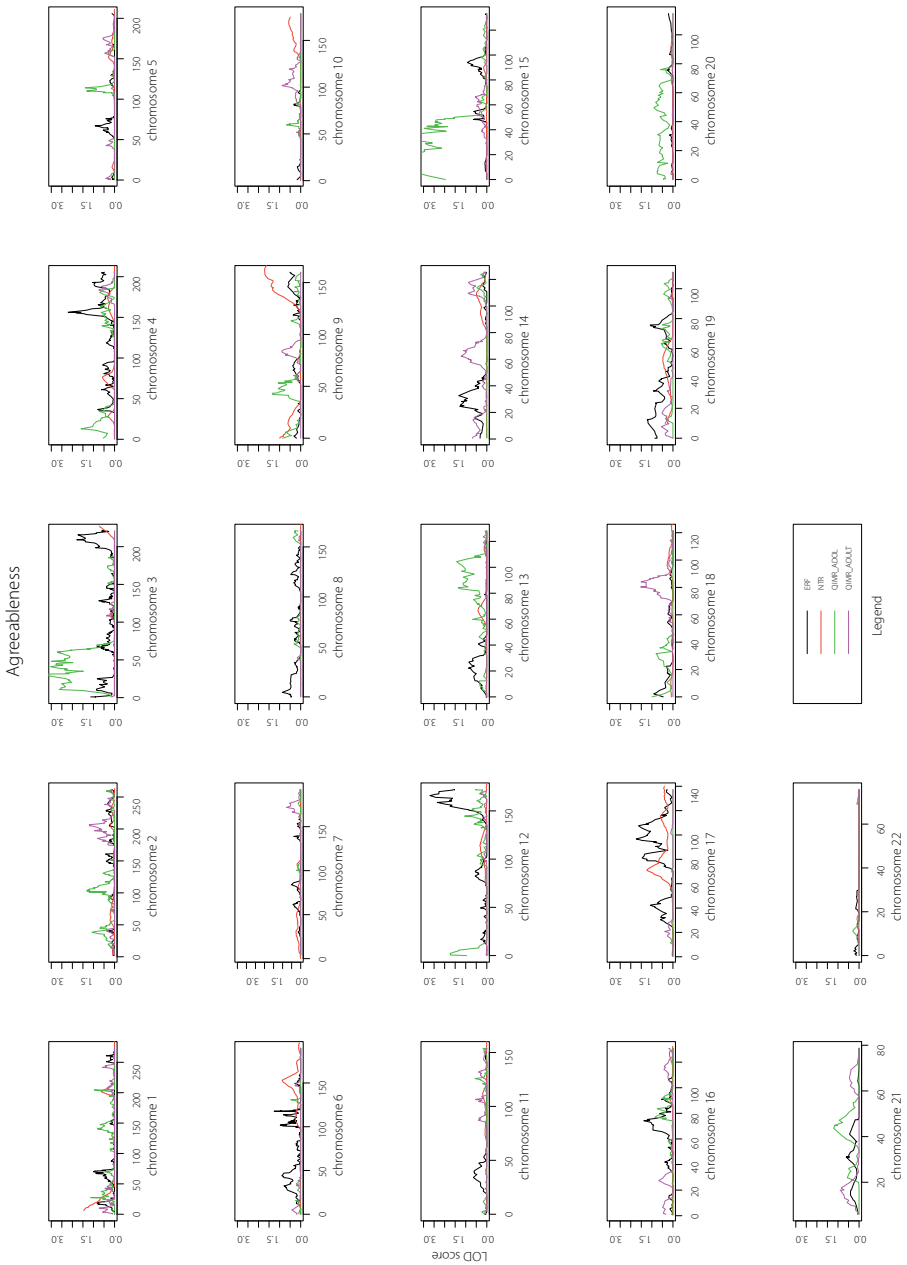
Supplementary Figure 2. Genome-wide linkage plots for individual scans for extraversion. The vertical axis shows the log of odds score and the horizontal axis shows the distance on the chromosome in centiMorgan.



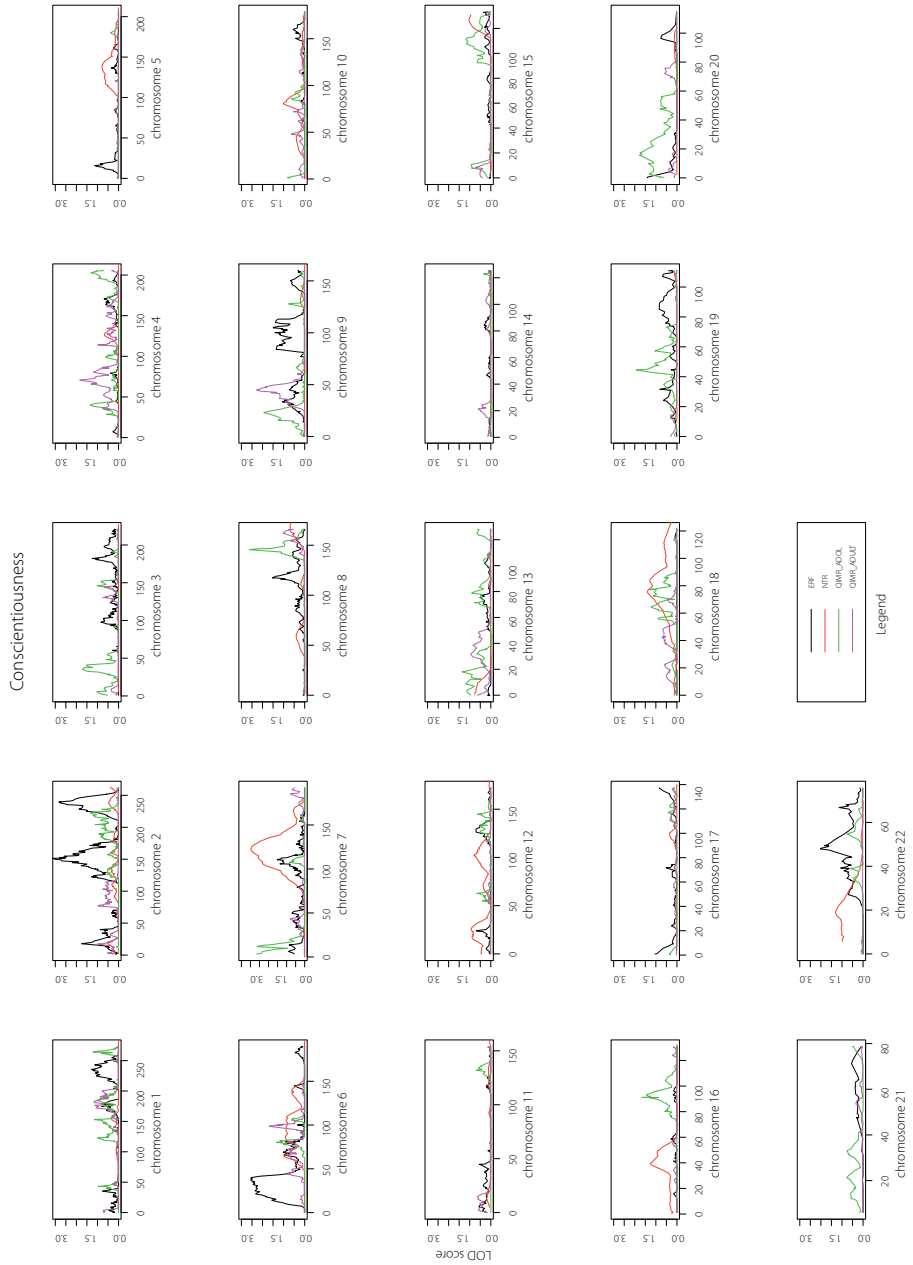
Supplementary Figure 3. Genome-wide linkage plots for individual scans for openness. The vertical axis shows the log of odds score and the horizontal axis shows the distance on the chromosome in centiMorgan.



Supplementary Figure 4. Genome-wide linkage plots for individual scans for agreeableness. The vertical axis shows the log of odds score and the horizontal axis shows the distance on the chromosome in centiMorgan.



Supplementary Figure 5. Genome-wide linkage plots for individual scans for conscientiousness. The vertical axis shows the log of odds score and the horizontal axis shows the distance on the chromosome in centiMorgan.



Chapter 3.3

Maternally transmitted *SCD5* is associated with agreeableness

Najaf Amin[1], A. Cecile J. W. Janssens[2], Narelle K. Hansell[3], Sarah E. Medland[3], Yurii S. Aulchenko[1], Gu Zhu[3], Nicholas G. Martin[3], Ben A. Oostra[1,4], Cornelia M. van Duijn[1†]

[1]Genetic Epidemiology Unit, Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands
[2]Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands
[3]Queensland Institute of Medical Research, Brisbane, QLD, Australia
[4]Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, the Netherlands

Chapter 4

Genome-wide association studies of Sleep traits

Chapter 4.1

Common genetic variant in *HRNBP3* is associated with sleep latency

Najaf Amin,[1] Karla V. Allebrandt,[2] Bertram Müller-Myhsok,[3] Karin Hek,[4,5] Maris Teder-Laving,[6] Caroline Hayward,[7] Tõnu Esko,[6] Josine G. van Mill,[8] Scott A. Melville,[9] Nicole Vogelzangs,[8] Harry Campbell,[10] Ayse Demirkan,[1] Zoran Dogas,[11] Fran Borovecki,[12] Jennifer E.Huffman,[7] Yurii S. Aulchenko,[1] Andre G. Uitterlinden,[13] Eric Wichmann,[14] Albert Hofman,[4] Ben A. Oostra,[1] Alan F. Wright,[7] James F. Wilson,[10] Peter P. Pramstaller,[9,15,16] Andrew A. Hicks,[9] A. Cecile J.W. Janssens,[4] Martha Merrow,[17] Igor Rudan,[18] Henning Tiemeier,[4] Brenda W. Penninx,[8] Andres Metspalu,[6] Thomas Meitinger,[19] Till Roenneberg,[2+] and Cornelia M. van Duijn,[1,20*+]

[1] Unit of Genetic Epidemiology, Department of Epidemiology, Erasmus University Medical Centre, Rotterdam, the Netherlands; [2] Institute of Medical Psychology, Ludwig-Maximilians-University, Munich, Germany; [3] Max-Planck-Institute of Psychiatry, Munich, Germany; [4] Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands; [5] O3 Research Center, Department of Psychiatry, Erasmus Medical Center, Rotterdam, The Netherlands; [6] Estonian Genome Project, University of Tartu and Estonian Biocentre, Tartu, Estonia; [7] Medical Research Council, Human Genetics Unit, IGMM, Edinburgh, Scotland; [8] Department of Psychiatry, VU University Medical Center Amsterdam, Amsterdam, The Netherlands; [9] Institute of Genetic Medicine, European Academy of Bolzano, Bolzano, Italy - affiliated institute of the University of Lübeck, Germany; [10] Centre For Population Health Sciences, University of Edinburgh, Edinburgh, Scotland; [11] Department of Neuroscience, Faculty of Medicine, University of Split, Split, Croatia; [12] Centre for Functional Genomics and Department of Neurology, Faculty of Medicine, University of Zagreb, Zagreb, Croatia; [13] Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands; [14] Institute of Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany; [15] Department of Neurology, General Central Hospital, Bolzano, Italy; [16] Department of Neurology, University of Lübeck, Lübeck, Germany; [17] Department of Chronobiology, University of Groningen, Haren, The Netherlands; [18] Croatian Centre for Global Health, University of Split Medical School, Split, Croatia; [19] Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany; [20] Centre of Medical Systems Biology, Netherlands Consortium on Health Aging and National Genomics Initiative

+ Authors contributed equally

Chapter 4.2

ABCC9 modulates sleep duration in humans and in *Drosophila*

Karla V. Allebrandt[1], Najaf Amin[2], Bertram Müller-Myhsok[3], Tõnu Esko[4], Maris Teder-Laving[4], Renata V.D.M. Azevedo[5], Caroline Hayward[6], Josine van Mill[7], Nicole Vogelzangs[7], Edward W. Green[5], Scott A. Melville[8], H.-Erich Wichmann[9], Ben A. Oostra[2], A. Cecile J.W. Janssens[2], Harry Campbell[10], James F. Wilson[10], Andrew A. Hicks[8], Peter P. Pramstaller[8], Zoran Dogas[11], Igor Rudan[10,11], Martha Merrow[12], Brenda Penninx[7], Charalambos P. Kyriacou[4], Andres Metspalu[5], Cornelia M. van Duijn[2,13], Thomas Meitinger[14] and Till Roenneberg[1*].

[1]Institute of Medical Psychology, Ludwig-Maximilians-University, Munich, Germany;
[2]Genetic Epidemiology Unit, Department of Epidemiology and Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands;
[3]Max-Planck-Institute of Psychiatry, Munich, Germany;
[4]Estonian Genome Center and Institute of Molecular and Cell Biology of University of Tartu, Estonian Biocentre, Tartu, Estonia;
[5]Department of Genetics, University of Leicester, Leicester, UK;
[6]Medical Research Council, Human Genetics Unit, IGMM, Edinburgh, Scotland;
[7]VU University Medical Center Amsterdam, Amsterdam, The Netherlands;
[8]Institute of Genetic Medicine, European Academy of Bolzano, Bozen, Italy – affiliated Institute of the University of Lübeck, Germany;
[9]Institute of Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany - Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-University, Munich, Germany, and Klinikum Großhadern, Munich Germany;
[10]Centre for Population Health Sciences, University of Edinburgh, Community Health Sciences, Edinburgh, Scotland;
[11]Croatian Centre for Global Health, University of Split Medical School, Split, Croatia;
[12]Department of Chronobiology, University of Groningen, Haren, The Netherlands;
[13]Centre of Medical Systems Biology, Netherlands Genomics Initiative (NGI), the Netherlands;
[14]Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany.

Chapter 4.3

Genome-wide association and gene expression analyses suggest association of *CYP1A1/CYP1A2* with coffee drinking

Najaf Amin[1]*, Enda Byrne[2]*, Julie Johnson[2], Georgia Chenevix-Trench[2], Stefan Walter[3,4], Ilja M. Nolte[5], kConFab Investigators[6], Jacqueline M. Vink[7], Rajesh Rawal[8], Massimo Mangino[9], Alexander Teumer[10], Joost C. Keers[11], Germaine Verwoert[3], Sebastian Baumeister[12], Reiner Biffar[13], Astrid Petersmann[14], Norbert Dahmen[15], Angela Doering[8], Aaron J. Isaacs[1], Linda Broer[1], Naomi R. Wray[2], Grant W. Montgomery[2], Daniel Levy[16,17], Bruce M. Psaty[18,19], Vilmundur Gudnason[20,21], Aravinda Chakravarti[22,23], Unnur Thorsteinsdottir[24], Frank J.A. van Rooij[4], Yurii S. Aulchenko[1], Jouke J. Hottenga[7], Fernando R. Rivadeneira[25], Albert Hofman[3], Andre G. Uitterlinden[25], Chris J. Hammond[26], So-Youn Shin[26], Arfan Ikram[3], Jacqueline C.M. Witteman[3], A. Cecile J.W. Janssens[3], Harold Snieder[5,11], Henning Tiemier[3,27], Bruce H. R. Wolfenbuttel[11,28], Ben A. Oostra[1,29], Andrew C. Heath[30], Erich Wichmann[8,31], Tim D. Spector[9], Hans J. Grabe[32], Dorret I. Boomsma[7], Nicholas G. Martin[2]*, Cornelia M. van Duijn[1,33]*†

[1]Unit of Genetic Epidemiology, Department of Epidemiology, Erasmus University Medical Center, the Netherlands
[2]Queensland Institute of Medical Research, Brisbane, Australia
[3]Department of Epidemiology, Erasmus University Medical Center, the Netherlands
[4]Department of Public Health, Erasmus University Medical Center, the Netherlands
[5]Unit of Genetic Epidemiology & Bioinformatics, Department of Epidemiology University Medical Center Groningen, University of Groningen, The Netherlands
[6]Peter MacCallum Cancer Institute, Melbourne, Australia
[7]Department of Biological Psychology, VU University Amsterdam, Amsterdam, the Netherlands
[8]Institute of Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany
[9]Department of Twin Research and Genetic Epidemiology, St. Thomas' Hospital Campus, King's College London, London, UK
[10]Interfaculty Institute for Genetics and Functional Genomics, University of Greifswald, Greifswald, Germany

[11]LifeLines Cohort Study & Biobank, University Medical Center Groningen, University of Groningen, the Netherlands
[12]Institute for Community Medicine, University of Greifswald, Germany
[13]Department of Prosthodontics, Gerodontology and Dental Materials, Center of Oral Health, University of Greifswald, Germany
[14]Institute of Clinical Chemistry and Laboratory Medicine, University of Greifswald, Germany
[15]Department of Psychiatry, University of Mainz, Mainz, Germany
[16]National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA
[17]Center for Population Studies, NHLBI, Bethesda, Maryland, USA
[18]Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA, USA
[19]Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA
[20]Icelandic Heart Association, Kopavogur, Iceland
[21]University of Iceland, Reykjavik, Iceland
[22]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland, USA
[23]Department of Epidemiology and Medicine, Johns Hopkins University, Baltimore, Maryland, USA
[24]deCODE Genetics, Reykjavik, Iceland
[25]Department of internal medicine, Erasmus University Medical Center, Rotterdam, the Netherlands
[26]Human Genetics, Wellcome Trust Sanger Institute, Genome Campus, Hinxton, UK
[27]Department of Child and Adolescent Psychiatry, Erasmus University Medical Center, Rotterdam, The Netherlands
[28]Department of Endocrinology, University Medical Center Groningen, University of Groningen, the Netherlands
[29]Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, the Netherlands
[30]Department of Psychiatry, Washington University, St Louis, USA
[31]Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany
[32]Department of Psychiatry and Psychotherapy, University of Greifswald, Stralsund, Germany
[33]Centre of Medical Systems Biology, Netherlands Consortium on Health Aging and National Genomics Initiative

* Authors contributed equally

ABSTRACT

Coffee consumption is a model for addictive behaviour. We performed a meta-analysis of genome-wide association scans (GWAS) on coffee intake from eight Caucasian cohorts ($N >$ 18000) and sought replication of our top findings in a further 8000 individuals. We also performed a gene expression analysis treating different cell lines with caffeine in order to identify caffeine-responsive genes that would be good candidates for genes influencing coffee-drinking habits Genome-wide significant association was observed for two single nucleotide polymorphisms (SNPs) in the 15q24 region. The two SNPs rs2470893 and rs2472297 (p-values $=1.6*10^{-11}$ & $2.7*10^{-11}$), which were also in strong linkage disequilibrium ($r^2 > 0.6$) with each other lie in the 23-kilobases long commonly shared 5-prime flanking region between *CYP1A1* and *CYP1A2* genes. *CYP1A1* was found to be down-regulated in the lymphoblastoid cell lines treated with caffeine. *CYP1A1* is known to metabolize polycyclic aromatic hydrocarbons which are important constituents of coffee, while *CYP1A2* is involved in primary metabolism of caffeine. Significant evidence of association was also detected at rs382140 (p-value = $3.9*10^{-09}$) near the *NRCAM* -- a gene implicated in vulnerability to addiction, and at another independent hit rs6495122 (p-value = $7.1*10^{-09}$) – a SNP associated with blood pressure -- in the 15q24 region near gene *ULK3*, in the meta-analysis of discovery and replication cohorts. Our results from GWAS and expression analysis also strongly implicate *CAB39L* as a gene that influences coffee drinking. Moreover, association analysis of our top hits with coffee-related phenotypes suggests that the 'T' allele of rs2470893, which is positively associated with coffee drinking, is also positively associated with systolic and diastolic blood pressure.

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 and type 2 diabetes[4]. However, coffee intake has also been associated with increased risk of some cancers[5-7], blood pressure[8-9] and myocardial infarction[10]. 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[11-12]. However, roasting of coffee beans is known to produce polycyclic aromatic hydrocarbons (PAH)[13] which are a class of carcinogenic chemicals that are formed by the incomplete combustion of organic matter. At unusually high doses, coffee is known to have potentiating effects on mutagenesis[12] including cytotoxicity of X-rays, ultraviolet light and chemotherapeutic agents[12]. Most of these mutagenic effects are known to be independent of caffeine[14] and have been attributed to aliphatic dicarbonyls[14-16] and hydrogen peroxide[15].

The P450 system in the liver plays a key role in coffee metabolism. The cytochrome P4501A1 encoded by the gene *CYP1A1* is known to metabolize PAHs such as benzo(a)pyrene. The caffeine content (~100mg/cup of coffee) is primarily metabolised in the liver by the cytochrome p450 *CYP1A2*, and is further broken down by the enzymes *CYP2A6* and *NAT2*[17-20]. Most of the biological effects of caffeine including the effects on the brain and central nervous system are mediated through antagonism of the adenosine receptors, specifically the A1 and A2A receptors[21]. Accumulating evidence from a number of sources points to the A_{2A} receptor as the main target for caffeine[22]. Under normal conditions, adenosine is hypothesised to activate the adenosine receptors, leading to subsequent activation of adenylyl cyclase and of Ca^{2+} channels[21]. 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 receptors[23], and changes in gene expression. It is through these mechanisms that caffeine mediates its effects on the brain and behaviour[21]. These changes in gene expression are not well understood, but the use of microarrays may give insight into the molecular changes brought about by caffeine[24].

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[25] and withdrawal effects when coffee consumption is stopped have been discussed. Many users experience withdrawal symptoms which include headache, decreased alertness and concentration, and depressed mood and irritability[26-27]. Beneficial effects from caffeine include improved psychomotor speed[28], mood[29] and alertness[30]. Several studies have shown that subjects reported higher levels of alertness

and concentration along with increased appetite for work[21; 29]. Some have postulated that the positive effects can be explained by a reversal of the withdrawal symptoms in habitual users[31-32], but this has been refuted by studies of caffeine intake in non-habitual users[33-34]. It is also clear that some users experience negative effects such as insomnia, anxiety and dysphoria[35-36]. Caffeine consumption is also known to play a role in suicide[37].

Genetic studies in twins suggest that the heritability estimates of coffee consumption range from 0.39 to 0.56[30,38-39]. Most genetic studies have focussed 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[40-41], due to both genetic and environmental factors[42]. There is some evidence, though not genome-wide significant, that polymorphisms in the gene are known to moderate the association between coffee consumption and hypertension [43] and myocardial infarction[10] as well as the risk breast cancer risk in BRCA1 carriers[44]. No association has been found between variants in *CYP1A2* and caffeine consumption[22], but a SNP in this gene (rs762551) has been shown to be associated with high inducibility of the *CYP1A2* enzyme in smokers[45]. A candidate gene study associated a polymorphism in the *ADORA2A* gene (rs5751876) with caffeine consumption[22] in a Costa Rican sample. This SNP has also been implicated in increasing risk to panic disorder in two separate studies in Caucasian populations[46-47] but none of the findings on *ADORA2A* reached genome wide significance nor were they replicated.

We have conducted a genome-wide association analysis (GWAS) meta-analysis of >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 whose pattern of expression is changed after caffeine treatment and which harbor polymorphisms that show evidence of association.

MATERIAL AND METHODS

Study populations

The study included participants from eight cohorts including

Erasmus Rucphen Family (ERF): The Erasmus Rucphen Family study is a family based study that includes over 3000 participants descending from 22 couples living in the Rucphen region in the 19th century. All living descendants of these couples and their spouses were invited to take part in the study. The frequency of coffee consumption and frequency was assessed with a questionnaire. 1814 participants who had both phenotype and genome-wide genotype data (54% women)[48] 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, 10/1999 – 7/2001). It was conducted between October 2006 and May 2008. From the KORA F4 survey (full cohort *n* = 3080), 1814 individuals aged between 32 to 81 years were selected for genotyping on the Affymetrix 1000K[49] chip. Coffee consumption was assessed with a questionnaire asking the number of cups of coffee consumed per day.

Rotterdam Study (RS-I & RS-II): The Rotterdam Study-I (RS-I) is a prospective population-based cohort study of 7,983 residents aged 55 years and older 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[50]. The Rotterdam Study-II (RS-II) is a prospective population-based cohort study of respectively 3,011 residents aged 55 years and older and the coffee consumption was assessed in the same manner.

Netherlands Twin Register (NTR): A sample of (mostly) adult twins was obtained from the Netherlands Twin Register (NTR), which was established in 1987 and contains information about Dutch twins and their families voluntarily taking part in research[51]. Since 1991, every 2 to 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 send out[52] 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 (SD 10.9). Genome-wide genotyping was available for 1087 subjects with coffee data.

The Study of Health in Pomerania (SHIP): SHIP is a longitudinal population-based cohort study in West Pomerania, a region in the northeast of. The baseline sample SHIP-0 comprised 4308 subjects[53-54]. 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 (www.twinsUK.ac.uk). These unselected twins were recruited from the general population through national media campaigns in the UK and shown to be comparable to age-matched population singletons in terms of disease-related and lifestyle characteristics [55-56].The TwinsUK I and II cohorts consist of twins from the adult twin British registry, also shown to be representative of singleton populations and the United Kingdom population[55]. Coffee consumption was assessed by questionnaire and genome-wide genotype and imputed data were available for 1,092 (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 to the study.

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?"[27]. The age range of respondents was 17-88 with a mean age of 31. 1,988 unrelated individuals provided both phenotype and genotype information. Detailed descriptions of phenotyping, genotyping, imputation and QC protocol are given elsewhere[27; 57].

Genotyping & Imputation

The participating cohorts were genotyped on commercially available platforms including Affymetrix, Illumina and Perlegen (**Supplementary Table 1**). Quality control was done in each group separately. The overall criteria were to exclude individuals with low call rate, excess heterozygosity, and gender mismatch. Based on 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[58-59] or IMPUTE[60-61].

Genome-wide association and Meta analysis

We conducted a meta-analysis of 10 GWAS from eight cohorts, consisting of more than 18,000 individuals and about 2.6 million imputed and genotyped SNPs. For each GWAS the association analysis was performed using linear regression analysis by regressing coffee categories[62] on age, sex and SNP allele dosage in ProbABEL[63], SNPTEST[64] or QUICKTEST (http://toby.freeshell.org/software/quicktest.shtml) (**Supplementary Table 1**). The coffee categories were defined as (1) 0-2 cups/day, (2) 3-4 cups/day, (3) 5-6 cups/day, (4) 7-9 cups/day and (5) 10 or more cups/day[62]. For ERF, which is a family-based cohort, the analysis was done using a mixed model[65-66] incorporating a relationship matrix estimated from the genotyped data[67]. A fixed effects meta-analysis was conducted in METAL using the inverse variance weighted method. All SNPs that had a low minor allele frequency (MAF < 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 prior to the meta-analysis.

Replication Analysis

We sought *in-silico* replication of the SNPs that showed a p-value < $1*10^{-06}$ (**Table 1**) in an independent Dutch cohort (LifeLines, N = 7929, % of women = 57). The LifeLines Cohort Study is a multi-disciplinary prospective population-based cohort study examining in a unique three-generation design the health and health-related behaviours of 165,000 persons living in the North East region of The Netherlands. All survey participants are between 18 and 90 years old at the time of enrollment. Recruitment has been going on since the end of 2006, and until January 2011 over 40,000 participants have been included. The genotyping for LifeLines was performed on Illumina CytoSNP12 v2. Genetic Imputations were performed in BEAGLE v3.1.0 using build 36 HapMap CEU v22 as the reference population. The statistical analysis was performed in PLINK using the same analysis model as in the discovery phase.

Gene Expression Analysis

We used public gene expression databases for lymphocytes and brain to examine the expression of genes implicated by the GWAS[68-70] (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 (GINI)[71]: 1) lymphoblastoid cell lines (LCLs) established from the blood of 24 females from hereditary non-*BRCA1/2* breast cancer families recruited into the Kathleen Cuningham Foundation for Research into Breast Cancer (kConFab), 2) a cell line newly established from a breast to bone metastasis; 3) the colon cancer cell line, HT29. Experiments were performed in triplicate for each sample. Optimisation experiments determined cell number and caffeine concentration for each cell type. For LCLs, 3.5 x 10^6 cells were plated in 10mls of tissue culture medium (RPMI-1640 supplemented with 10% fetal calf serum) containing 7.5mM caffeine (Sigma). For both the HT29 and bone metastasis cell line, 1 x 10^6 cells were plated in 10mls of tissue culture medium (RPMI-1640 supplemented with 10% fetal calf serum) 24 hours prior to treating the cells with media containing 10mM caffeine (Sigma). All cell lines were incubated with caffeine for a total of eight hours 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) as per the manufacturer's instructions. Biotinylated cRNA was prepared from 450ng of total RNA using the Illumina TotalPrep RNA Amplification Kit (Ambion). After quantification using a ND-1000 spectrophotometer (Nanodrop Technologies), a total of 156 samples (750ng cRNA per sample) were hybridised to HumanHT-12 Expression BeadChips (Illumina) using all manufacturer's reagents for washing, detecting and scanning as per the Whole-Genome Gene Expression Direct Hybridisation Assay protocol (Illumina).

The HumanHT-12 Expression BeadChips (Illumina) contains 48,803 probes that cover more than 25,000 annotated genes. Expression data were collated and quality checked in Illumina BeadStudio and then imported into GeneSpring V10.0 (Agilent Technologies). Data were quantile normalised 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, 2005) 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 (B statistic)> 0 was selected. Then, a hierarchical search of genes based on log Fold Changes. Gene lists were then imported into GeneSpring V10.0 (Agilent Technologies) for data visualisation and to determine overlapping genes between different cell types.

RESULTS

Genome-wide Association analysis

The descriptive statistics of each cohort are provided in the **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 ($p = 2.3*10^{-08}$). 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**). Neither of the SNPs are in strong linkage disequilibrium (LD) with rs762551, the SNP previously identified as increasing *CYP1A2* inducibility in caffeine in smokers ($r^2 = 0.12$ and 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. Rs2470893 was genotyped in four cohorts and in these the minor allele frequency ranged from 0.26-0.35. 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 = $8.5*10^{-04}$; rs2472297, p-value = $8.8*10^{-04}$) and RS-II (rs2470893, p-value = $5.3*10^{-07}$; rs2472297, p-value = $6.3*10^{-08}$), which were the primary cohorts driving the association signal at 15q24.

Table 1. Top hits of the meta-analysis of the GWAS

MarkerName	Effect allele	Non-effect allele	β	σ_β	p-value	Direction	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]

Figure 1 shows that there are a number of signals in other genes in the chromosome 15 region. Strong association was observed for rs6495122 ($p = 8.22 \times 10^{-08}$) (**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 the RS-II (which was the largest contributor to the association signal at 15q24, p-value = $8.4*10^{-06}$) adjusting for the two most significant findings (rs2470893 and rs2472297) revealed strong association of rs6495122 (p-value = $2.6*10^{-07}$, $\beta = 0.19$) and rs12487 (p-value = $1.9*10^{-07}$, $\beta = 0.19$) with coffee drinking (**Supplementary Figure 3**).

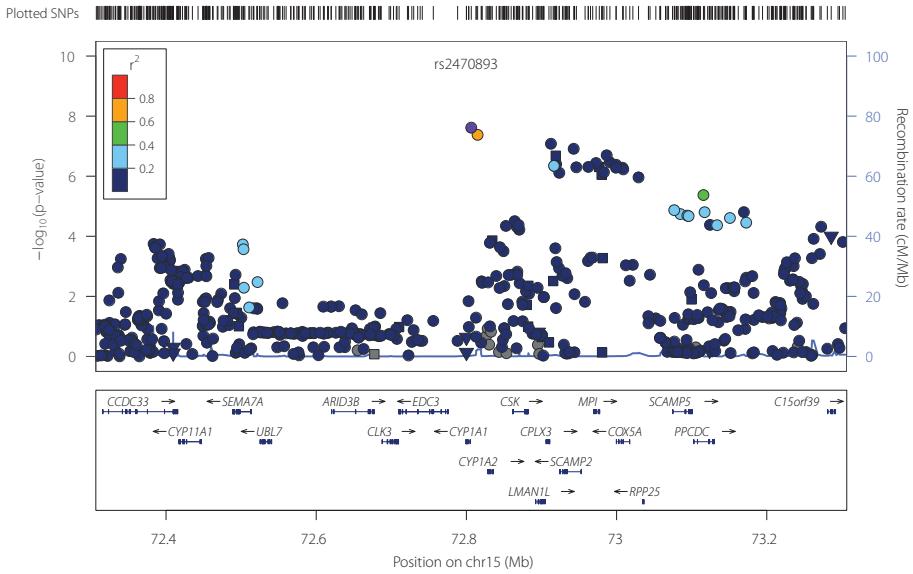


Figure 1. Regional association plot for 15q24. The vertical axis shows the negative logarithm of the association p-values and the horizontal axis shows the position in mega bases. Each dot represents a SNP and the colours of the dots represent the extent of linkage disequilibrium with the top SNP. Genes in the region are shown below the horizontal axis.

Other SNPs on different chromosomes that showed strong evidence of association (p-values below $1*10^{-06}$) 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**).

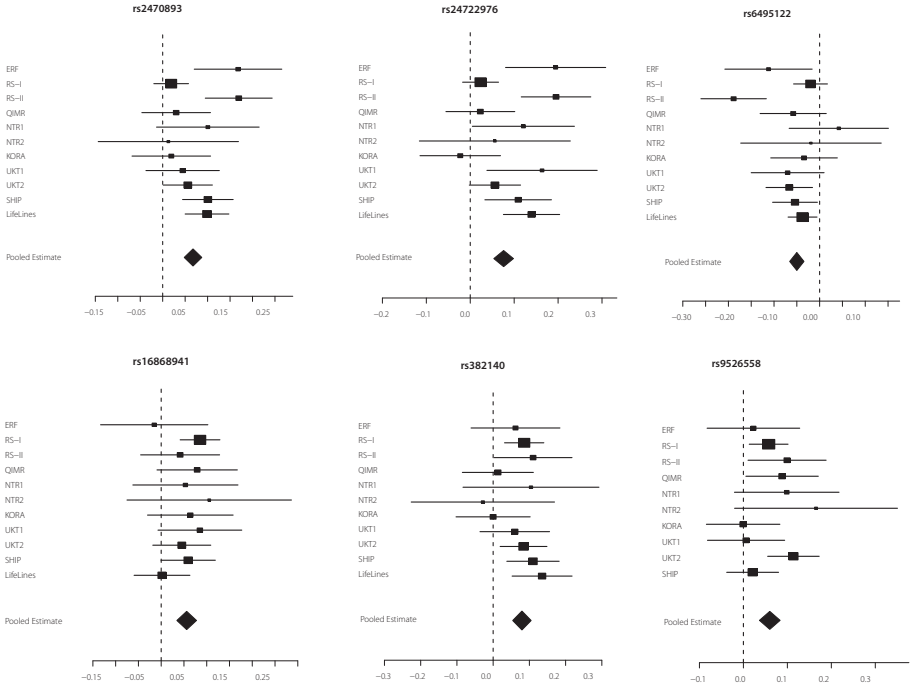


Figure 2. Forest plot for the top hits from the GWAS. Vertically left, the populations included in the GWAS and the replication phase. The boxes represent the precision and horizontal lines represent the confidence intervals. The diamond represents the pooled effect estimate from the meta-analysis of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

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 the **Table 1** which necessitated a Bonferroni corrected significance threshold of $8.3*10^{-03}$. Because of bad imputation quality in the replication cohort, no replication was performed for rs9526558. 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 rs6495122 with coffee drinking (p-value = $7.1*10^{-09}$) (**Table 2, Figure 2**).

Table 2. Results of replication analysis and meta-analysis of discovery and replication cohorts

Marker-Name	Effect Allele	Non-effect Allele	chr	position	gene	Discovery (*N*= 18,000)			Replication (*N* = 7,929)			Meta-analysis (Discovery + replication)		
						β	σβ	p-value	β	σβ	p-value	β	σβ	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

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 brain[70]. **Supplementary table 3** shows that rs2470893 is involved in the expression of the *COX5A* gene (p-value = $1.2*10^{-04}$). Also 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 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 2 diabetes mellitus is suggested to be at least partially, explained by the iron absorption inhibitory effect of coffee[72-73].

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 PAHs was found to be down-regulated by caffeine in LCLs (log-fold-change (logFC) = 0.109, p-value$_{adj}$ = $5.38*10^{-08}$, β = 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 this 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 following caffeine treatment was most pronounced in LCLs (logFC = -0.64, p-value$_{adj}$ = $3.8*10^{-27}$, β = 55.03). In combination with the GWAS results, with finding, this finding suggests that *CAB39L* is involved 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.

Coffee related phenotypes

We tested for the association of the top hits from the GWAS with two established coffee related phenotypes: blood pressure and Alzheimer's disease. For blood pressure and hypertension we used the data from Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE). Significant association was found for the top hit rs2470893 with systolic (p-value = $5*10^{-03}$) and diastolic blood pressure (p-value = $1.6*10^{-04}$) and a borderline association was observed with hypertension (p-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[74]. For Alzheimer Disease (AD), we used the data from Rotterdam Study (RS) and two previously published GWAS on AD[75-76]. No association of any of the SNPs tested was detected with AD.

DISCUSSION

We conducted a meta-analysis of GWAS on coffee consumption from eight 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 lymphoblastoid cell lines strongly implicates the*CYP1A1/CYP1A2* locus in coffee drinking. Our study also suggests significant association of a SNP rs382140 in the promoter region of the *NRCAM* gene with coffee drinking, and the combined results from association (rs9526558) and expression analysis implicate *CAB39L* in coffee drinking.

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[77]. The *CYP1A2* gene encodes a P450 enzyme involved in O-de-ethylation of phenacetin[78]. The human hepatic microsomal caffeine 3-demethylation, which is the initial major step in caffeine biotransformation in humans, is selectively catalyzed by *CYP1A2* [78]. *CYP1A1* encodes the member P1-450 of the cytochrome P450 superfamily of enzymes. P1-450 is most closely associated with polycyclic-hydrocarbon-induced AHH activity[67]and known to metabolize PAHs such as benzo(a) pyrene, which is a constituent of coffee known to be involved in carcinogenesis[12]. 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-prime flanking region[79]. 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 human brain[80] and localized to the cortical regions, midbrain, basal ganglia and cerebellum[81], while *CYP1A2* has been found in most brain regions examined[82-83].

A third SNP - rs6495122 – approached genome-wide significance ($p = 8.22 \times 10^{-08}$). 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 0.086 respectively). Also the secondary conditional analysis in RS-II suggests that rs6495122 is independent of the two most significant hits in this region (rs2470893 and rs2472297) (**Supplementary Figure 3**). A previous meta-analysis found this SNP to be significantly associated with diastolic blood pressure, and nominally associated with systolic blood pressure and hypertension[74]. There is evidence that caffeine consumption is associated with an increase in blood pressure. Consuming a dose of caffeine equivalent to 2-3 cups of coffee (200-250mg) has been found to increase both systolic and diastolic blood pressure[9]. Results from studies examining the long-term effects of caffeine on blood pressure and risk to hypertension have been mixed[20]. 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. *NRCAM* encodes Neuronal cell adhesion molecule, is expressed in brain and is involved in several aspects of nervous system development. Allelic variants of this gene have been associated with autism[84] and addiction vulnerability[85-86]. Genetic variants that increase 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[87-88]. Caffeine has been described as "a model drug of abuse"[25; 89-90] and the finding of the *NRCAM* gene are in line with this hypothesis. Expression of *NRCAM* was also found to be up regulated in papillary thyroid carcinomas[91].

CAB39L was found to be up-regulated 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 = $6.8*10^{-07}$). *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= $1.5*10^{-07}$). Like *CAB39L*, *NCALD* is involved in calcium metabolism. *NCALD* encodes a member of the neuronal calcium sensor (NCS) 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 thought to be 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[92]. 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[93] thereby delaying the onset of sleep.

In this study, we have combined gene expression data from three cell types treated with caffeine with GWAS data to study gene implicated in coffee drinking habits in more detail. Gene expression in lymphocytes strengthens the evidence of association for *CYP1A1* and *CAB39L*. Although in analysing the effects of caffeine, it would be most pertinent to study changes in other cells and tissues including liver and brain tissue, 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[94]. 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[24], 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 to controls may provide more power for detecting caffeine sensitivity 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.

In summary, we report association of the *CYP1A1/CYP1A2* locus with coffee drinking. The region is also significantly implicated in a separate meta-analysis of GWAS of coffee drinking conducted by the deCODE in about 7000 coffee drinkers (deCODE, unpublished data). We further report association of the variant rs382140 near the *NRCAM* gene and an intronic variant (rs9526558) in the *CAB39L* gene with coffee consumption. No evidence for the *NRCAM* gene was observed in the deCODE data, which may be attributed to the different study design used in the two studies (exclusion of coffee non-drinkers in deCODE). The evidence of association of the *CAB39L* gene is further supported by the fact that consistent effects (with respect to size and direction) were observed for the rs9526558 in the deCODE samples. These GWAS findings are further strengthened by differential expression of *CYP1A1* and *CAB39L* after caffeine treatment of different cell types.

AKCNOWLEDGEMENTS

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 Organisation 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.

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); 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. J.M.Vink is financially supported by NWO (VENI 451-06-004).

QIMR: We wish to thank Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow Up Study (funded 2001-2009 by NHMRC and currently by the National Breast Cancer Foundation and Cancer Australia #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 GenomEUtwin Project (QLG2-CT-2002-01254), and the U.S. 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). G.W.M and G.C.T are supported by the National Health and Medical Research Council (NHMRC) Fellowship Scheme. Julie Johnson was financially supported by the Cancer Council Queensland.

Rotterdam Study (RS-I & 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.

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 no. 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.

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 GenomEUtwin Project (QLG2-CT-2002-01254). The study also receives support from the Dept of Health via 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) .The authors acknowledge the funding and support of the National Eye Institute via 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, North Carolina, 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, Mitra Valimohammadi for their help in creating the GWAS database, and Rob Bieringa, Joost Keers, René Oostergo, Rosalie Visser, Judith Vonk for their work related to data-collection and validation. The authors 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

1. Ruusunen, A., Lehto, S.M., Tolmunen, T., Mursu, J., Kaplan, G.A., and Voutilainen, S. (2010). 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 13, 1215-1220.
2. Sugiyama, K., Kuriyama, S., Akhter, M., Kakizaki, M., Nakaya, N., Ohmori-Matsuda, K., Shimazu, T., Nagai, M., Sugawara, Y., Hozawa, A., et al. (2010). Coffee consumption and mortality due to all causes, cardiovascular disease, and cancer in Japanese women. J Nutr 140, 1007-1013.
3. Choi, H.K., and Curhan, G. (2010). Coffee consumption and risk of incident gout in women: the Nurses' Health Study. Am J Clin Nutr 92, 922-927.
4. van Dieren, S., Uiterwaal, C.S., van der Schouw, Y.T., van der, A.D., Boer, J.M., Spijkerman, A., Grobbee, D.E., and Beulens, J.W. (2009). Coffee and tea consumption and risk of type 2 diabetes. Diabetologia.
5. Trichopoulos, D., Papapostolou, M., and Polychronopoulou, A. (1981). Coffee and ovarian cancer. Int J Cancer 28, 691-693.
6. MacMahon, B., Yen, S., Trichopoulos, D., Warren, K., and Nardi, G. (1981). Coffee and cancer of the pancreas. N Engl J Med 304, 630-633.
7. Silvera, S.A., Jain, M., Howe, G.R., Miller, A.B., and Rohan, T.E. (2007). Intake of coffee and tea and risk of ovarian cancer: a prospective cohort study. Nutr Cancer 58, 22-27.
8. Jee, S.H., He, J., Whelton, P.K., Suh, I., and Klag, M.J. (1999). The effect of chronic coffee drinking on blood pressure: a meta-analysis of controlled clinical trials. Hypertension 33, 647-652.
9. Nurminen, M.L., Niittynen, L., Korpela, R., and Vapaatalo, H. (1999). Coffee, caffeine and blood pressure: a critical review. Eur J Clin Nutr 53, 831-839.
10. Cornelis, M.C., El-Sohemy, A., Kabagambe, E.K., and Campos, H. (2006). Coffee, CYP1A2 genotype, and risk of myocardial infarction. JAMA 295, 1135-1141.
11. Kalthoff, S., Ehmer, U., Freiberg, N., Manns, M.P., and Strassburg, C.P. (2010). Coffee induces expression of glucuronosyltransferases by the aryl hydrocarbon receptor and Nrf2 in liver and stomach. Gastroenterology 139, 1699-1710, 1710 e1691-1692.
12. Nehlig, A., and Debry, G. (1994). Potential genotoxic, mutagenic and antimutagenic effects of coffee: a review. Mutat Res 317, 145-162.
13. Kruijf, N., Schouten, A., and Van der Stegen, G.H.D. (1987). Occurence of benzo[a]pyrene in roasted coffee, instant coffee and coffee brew. The Cafe Cacao 35, 151-154.
14. Nagao, M., Takahashi, Y., Yamanaka, H., and Sugimura, T. (1979). Mutagens in coffee and tea. Mutat Res 68, 101-106.
15. Aeschbacher, H.U., Wolleb, U., Loliger, J., Spadone, J.C., and Liardon, R. (1989). Contribution of coffee aroma constituents to the mutagenicity of coffee. Food Chem Toxicol 27, 227-232.
16. Nukaya, H., Iwami, T., Ishida, H., Tsuji, K., Suwa, Y., Wakabayashi, K., Nagao, M., Sugimura, T., and Kosuge, T. (1990). N-2 acetylation of 2'-deoxyguanosine by coffee mutagens, methylglyoxal and hydrogen peroxide. Mutat Res 245, 251-257.
17. Krul, C., and Hageman, G. (1998). Analysis of urinary caffeine metabolites to assess biotransformation enzyme activities by reversed-phase high-performance liquid chromatography. J Chromatogr B Biomed Sci Appl 709, 27-34.
18. Crews, H.M., Olivier, L., and Wilson, L.A. (2001). Urinary biomarkers for assessing dietary exposure to caffeine. Food Addit Contam 18, 1075-1087.
19. Selbach, O., and Haas, H.L. (2006). Hypocretins: the timing of sleep and waking. Chronobiol Int 23, 63-70.
20. Higdon, J.V., and Frei, B. (2006). Coffee and health: a review of recent human research. Crit Rev Food Sci Nutr 46, 101-123.
21. Fredholm, B.B., Battig, K., Holmen, J., Nehlig, A., and Zvartau, E.E. (1999). Actions of caffeine in the brain with special reference to factors that contribute to its widespread use. Pharmacol Rev 51, 83-133.

22. Cornelis, M.C., El-Sohemy, A., and Campos, H. (2007). Genetic polymorphism of the adenosine A2A receptor is associated with habitual caffeine consumption. Am J Clin Nutr 86, 240-244.

23. Ferre, S., von Euler, G., Johansson, B., Fredholm, B.B., and Fuxe, K. (1991). Stimulation of high-affinity adenosine A2 receptors decreases the affinity of dopamine D2 receptors in rat striatal membranes. Proc Natl Acad Sci U S A 88, 7238-7241.

24. Yu, L., Coelho, J.E., Zhang, X., Fu, Y., Tillman, A., Karaoz, U., Fredholm, B.B., Weng, Z., and Chen, J.F. (2009). Uncovering multiple molecular targets for caffeine using a drug target validation strategy combining A 2A receptor knockout mice with microarray profiling. Physiol Genomics 37, 199-210.

25. Holtzman, S.G. (1990). Caffeine as a model drug of abuse. Trends Pharmacol Sci 11, 355-356.

26. Juliano, L.M., and Griffiths, R.R. (2004). A critical review of caffeine withdrawal: empirical validation of symptoms and signs, incidence, severity, and associated features. Psychopharmacology (Berl) 176, 1-29.

27. Luciano, M., Kirk, K.M., Heath, A.C., and Martin, N.G. (2005). The genetics of tea and coffee drinking and preference for source of caffeine in a large community sample of Australian twins. Addiction 100, 1510-1517.

28. Mitchell, P.J., and Redman, J.R. (1992). Effects of caffeine, time of day and user history on study-related performance. Psychopharmacology (Berl) 109, 121-126.

29. Silverman, K., Mumford, G.K., and Griffiths, R.R. (1994). Enhancing caffeine reinforcement by behavioral requirements following drug ingestion. Psychopharmacology (Berl) 114, 424-432.

30. Horne, J.A., and Reyner, L.A. (1996). Counteracting driver sleepiness: effects of napping, caffeine, and placebo. Psychophysiology 33, 306-309.

31. James, J.E. (1994). Does caffeine enhance or merely restore degraded psychomotor performance? Neuropsychobiology 30, 124-125.

32. James, J.E. (1995). Caffeine and psychomotor performance revisited. Neuropsychobiology 31, 202-203.

33. Rogers, P.J., Richardson, N.J., and Dernoncourt, C. (1995). Caffeine use: is there a net benefit for mood and psychomotor performance? Neuropsychobiology 31, 195-199.

34. Christopher, G., Sutherland, D., and Smith, A. (2005). Effects of caffeine in non-withdrawn volunteers. Hum Psychopharmacol 20, 47-53.

35. Evans, S.M., and Griffiths, R.R. (1991). Dose-related caffeine discrimination in normal volunteers: individual differences in subjective effects and self-reported cues. Behav Pharmacol 2, 345-356.

36. Childs, E., Hohoff, C., Deckert, J., Xu, K., Badner, J., and de Wit, H. (2008). Association between ADORA2A and DRD2 polymorphisms and caffeine-induced anxiety. Neuropsychopharmacology 33, 2791-2800.

37. Klatsky, A.L., Armstrong, M.A., and Friedman, G.D. (1993). Coffee, tea, and mortality. Ann Epidemiol 3, 375-381.

38. 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.

39. Laitala, V.S., Kaprio, J., and Silventoinen, K. (2008). Genetics of coffee consumption and its stability. Addiction 103, 2054-2061.

40. Gu, L., Gonzalez, F.J., Kalow, W., and Tang, B.K. (1992). Biotransformation of caffeine, paraxanthine, theobromine and theophylline by cDNA-expressed human CYP1A2 and CYP2E1. Pharmacogenetics 2, 73-77.

41. Kalow, W., and Tang, B.K. (1993). The use of caffeine for enzyme assays: a critical appraisal. Clin Pharmacol Ther 53, 503-514.

42. Rasmussen, B.B., Brix, T.H., Kyvik, K.O., and Brosen, K. (2002). The interindividual differences in the 3-demthylation of caffeine alias CYP1A2 is determined by both genetic and environmental factors. Pharmacogenetics 12, 473-478.

43. Palatini, P., Ceolotto, G., Ragazzo, F., Dorigatti, F., Saladini, F., Papparella, I., Mos, L., Zanata, G., and Santonastaso, M. (2009). CYP1A2 genotype modifies the association between coffee intake and the risk of hypertension. J Hypertens 27, 1594-1601.

44. Kotsopoulos, J., Ghadirian, P., El-Sohemy, A., Lynch, H.T., Snyder, C., Daly, M., Domchek, S., Randall, S., Karlan, B., Zhang, P., et al. (2007). The CYP1A2 genotype modifies the association between coffee consumption and breast cancer risk among BRCA1 mutation carriers. Cancer Epidemiol Biomarkers Prev 16, 912-916.

45. 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.

46. Deckert, J., Nothen, M.M., Franke, P., Delmo, C., Fritze, J., Knapp, M., Maier, W., Beckmann, H., and Propping, P. (1998). 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 3, 81-85.

47. Hamilton, S.P., Slager, S.L., De Leon, A.B., Heiman, G.A., Klein, D.F., Hodge, S.E., Weissman, M.M., Fyer, A.J., and Knowles, J.A. (2004). Evidence for genetic linkage between a polymorphism in the adenosine 2A receptor and panic disorder. Neuropsychopharmacology 29, 558-565.

48. Aulchenko, Y.S., Heutink, P., Mackay, I., Bertoli-Avella, A.M., Pullen, J., Vaessen, N., Rademaker, T.A., Sandkuijl, L.A., Cardon, L., Oostra, B., et al. (2004). Linkage disequilibrium in young genetically isolated Dutch population. Eur J Hum Genet 12, 527-534.

49. Wichmann, H., Gieger, C., Illig, T., and Group, M. (2005). KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. . Gesundheitswesen 67 Suppl 1, S26-30.

50. Hofman, A., Breteler, M.M., van Duijn, C.M., Janssen, H.L., Krestin, G.P., Kuipers, E.J., Stricker, B.H., Tiemeier, H., Uitterlinden, A.G., Vingerling, J.R., et al. (2009). The Rotterdam Study: 2010 objectives and design update. Eur J Epidemiol 24, 553-572.

51. Boomsma, D.I., de Geus, E.J., Vink, J.M., Stubbe, J.H., Distel, M.A., Hottenga, J.J., Posthuma, D., van Beijsterveldt, T.C., Hudziak, J.J., Bartels, M., et al. (2006). Netherlands Twin Register: from twins to twin families. Twin Res Hum Genet 9, 849-857.

52. Vink, J.M., and Boomsma, D.I. (2008). A comparison of early and late respondents in a twin-family survey study. Twin Res Hum Genet 11, 165-173.

53. John, U., Greiner, B., Hensel, E., Ludemann, J., Piek, M., Sauer, S., Adam, C., Born, G., Alte, D., Greiser, E., et al. (2001). Study of Health In Pomerania (SHIP): a health examination survey in an east German region: objectives and design. Soz Praventivmed 46, 186-194.

54. Volzke, H., Alte, D., Schmidt, C.O., Radke, D., Lorbeer, R., Friedrich, N., Aumann, N., Lau, K., Piontek, M., Born, G., et al. (2010). Cohort Profile: The Study of Health in Pomerania. Int J Epidemiol.

55. Andrew, T., Hart, D.J., Snieder, H., de Lange, M., Spector, T.D., and MacGregor, A.J. (2001). Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. Twin Res 4, 464-477.

56. Spector, T.D., and Williams, F.M. (2006). The UK Adult Twin Registry (TwinsUK). Twin Res Hum Genet 9, 899-906.

57. Medland, S.E., Nyholt, D.R., Painter, J.N., McEvoy, B.P., McRae, A.F., Zhu, G., Gordon, S.D., Ferreira, M.A., Wright, M.J., Henders, A.K., et al. (2009). Common variants in the trichohyalin gene are associated with straight hair in Europeans. Am J Hum Genet 85, 750-755.

58. Li, Y., and Abecasis, G.R. (2006). Rapid Haplotype Reconstruction and Missing Genotype Inference. Am J Hum Genet. Am J Hum Genet S79, 2290.

59. Li, Y., Willer, C., Sanna, S., and Abecasis, G. (2009). Genotype imputation. Annu Rev Genomics Hum Genet 10, 387-406.

60. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 5, e1000529.

61. 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.

62. Tuomilehto, J., Hu, G., Bidel, S., Lindstrom, J., and Jousilahti, P. (2004). Coffee consumption and risk of type 2 diabetes mellitus among middle-aged Finnish men and women. JAMA 291, 1213-1219.

63. Aulchenko, Y.S., Struchalin, M.V., and van Duijn, C.M. (2010). ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics 11, 134.

64. Wellcome Trust Case Control, C. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447, 661-678.

65. Chen, W.M., and Abecasis, G.R. (2007). Family-based association tests for genomewide association scans. Am J Hum Genet 81, 913-926.

66. Aulchenko, Y.S., Ripke, S., Isaacs, A., and van Duijn, C.M. (2007). GenABEL: an R library for genome-wide association analysis. Bioinformatics 23, 1294-1296.

67. Amin, N., van Duijn, C.M., and Aulchenko, Y.S. (2007). A genomic background based method for association analysis in related individuals. PLoS One 2, e1274.

68. Stranger, B.E., Nica, A.C., Forrest, M.S., Dimas, A., Bird, C.P., Beazley, C., Ingle, C.E., Dunning, M., Flicek, P., Koller, D., et al. (2007). Population genomics of human gene expression. Nat Genet 39, 1217-1224.

69. Montgomery, S.B., Sammeth, M., Gutierrez-Arcelus, M., Lach, R.P., Ingle, C., Nisbett, J., Guigo, R., and Dermitzakis, E.T. (2010). Transcriptome genetics using second generation sequencing in a Caucasian population. Nature 464, 773-777.

70. Myers, A.J., Gibbs, J.R., Webster, J.A., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., Leung, D., Bryden, L., Nath, P., et al. (2007). A survey of genetic human cortical gene expression. Nat Genet 39, 1494-1499.

71. Ivanov, I., Lo, K.C., Hawthorn, L., Cowell, J.K., and Ionov, Y. (2007). Identifying candidate colon cancer tumor suppressor genes using inhibition of nonsense-mediated mRNA decay in colon cancer cells. Oncogene 26, 2873-2884.

72. Mascitelli, L., Pezzetta, F., and Sullivan, J.L. (2007). Inhibition of iron absorption by coffee and the reduced risk of type 2 diabetes mellitus. Arch Intern Med 167, 204-205; author reply 205.

73. Pereira, M.A., Parker, E.D., and Folsom, A.R. (2006). Coffee consumption and risk of type 2 diabetes mellitus: an 11-year prospective study of 28 812 postmenopausal women. Arch Intern Med 166, 1311-1316.

74. 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.

75. Lambert, J.C., Heath, S., Even, G., Campion, D., Sleegers, K., Hiltunen, M., Combarros, O., Zelenika, D., Bullido, M.J., Tavernier, B., et al. (2009). Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat Genet 41, 1094-1099.

76. Harold, D., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Hamshere, M.L., Pahwa, J.S., Moskvina, V., Dowzell, K., Williams, A., et al. (2009). Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat Genet 41, 1088-1093.

77. Castorena-Torres, F., Mendoza-Cantu, A., de Leon, M.B., Cisneros, B., Zapata-Perez, O., Lopez-Carrillo, L., Salinas, J.E., and Albores, A. (2005). CYP1A2 phenotype and genotype in a population from the Carboniferous Region of Coahuila, Mexico. Toxicol Lett 156, 331-339.

78. Butler, M.A., Iwasaki, M., Guengerich, F.P., and Kadlubar, F.F. (1989). 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 U S A 86, 7696-7700.

79. Corchero, J., Pimprale, S., Kimura, S., and Gonzalez, F.J. (2001). Organization of the CYP1A cluster on human chromosome 15: implications for gene regulation. Pharmacogenetics 11, 1-6.

80. Yun, C.H., Park, H.J., Kim, S.J., and Kim, H.K. (1998). Identification of cytochrome P450 1A1 in human brain. Biochem Biophys Res Commun 243, 808-810.

81. McFayden, M.C., Melvin, W.T., and Murray, G.I. (1998). Regional distribution of individual forms of cytochrome P450 mRNA in normal adult human brain. Biochem Pharmacol 55, 825-830.

82. Morse, D.C., Stein, A.P., Thomas, P.E., and Lowndes, H.E. (1998). Distribution and induction of cytochrome P450 1A1 and 1A2 in rat brain. Toxicol Appl Pharmacol 152, 232-239.

83. Farin, F.M., and Omiecinski, C.J. (1993). Regiospecific expression of cytochrome P-450s and microsomal epoxide hydrolase in human brain tissue. J Toxicol Environ Health 40, 317-335.

84. Sakurai, T., Ramoz, N., Reichert, J.G., Corwin, T.E., Kryzak, L., Smith, C.J., Silverman, J.M., Hollander, E., and Buxbaum, J.D. (2006). Association analysis of the NrCAM gene in autism and in subsets of families with severe obsessive-compulsive or self-stimulatory behaviors. Psychiatr Genet 16, 251-257.

85. Ishiguro, H., Liu, Q.R., Gong, J.P., Hall, F.S., Ujike, H., Morales, M., Sakurai, T., Grumet, M., and Uhl, G.R. (2006). NrCAM in addiction vulnerability: positional cloning, drug-regulation, haplotype-specific expression, and altered drug reward in knockout mice. Neuropsychopharmacology 31, 572-584.

86. Matzel, L.D., Babiarz, J., Townsend, D.A., Grossman, H.C., and Grumet, M. (2008). Neuronal cell adhesion molecule deletion induces a cognitive and behavioral phenotype reflective of impulsivity. Genes Brain Behav 7, 470-480.

87. Swan, G.E., Carmelli, D., Rosenman, R.H., Fabsitz, R.R., and Christian, J.C. (1990). Smoking and alcohol consumption in adult male twins: genetic heritability and shared environmental influences. J Subst Abuse 2, 39-50.

88. Swan, G.E., Carmelli, D., and Cardon, L.R. (1996). The consumption of tobacco, alcohol, and coffee in Caucasian male twins: a multivariate genetic analysis. J Subst Abuse 8, 19-31.

89. Strain, E.C., and Griffiths, R.R. (1995). Caffeine dependence: fact or fiction? J R Soc Med 88, 437-440.

90. Hughes, J.R., Oliveto, A.H., Liguori, A., Carpenter, J., and Howard, T. (1998). Endorsement of DSM-IV dependence criteria among caffeine users. Drug Alcohol Depend 52, 99-107.

91. Gorka, B., Skubis-Zegadlo, J., Mikula, M., Bardadin, K., Paliczka, E., and Czarnocka, B. (2007). NrCAM, a neuronal system cell-adhesion molecule, is induced in papillary thyroid carcinomas. Br J Cancer 97, 531-538.

92. Kamiyama, M., Kobayashi, M., Araki, S., Iida, A., Tsunoda, T., Kawai, K., Imanishi, M., Nomura, M., Babazono, T., Iwamoto, Y., et al. (2007). Polymorphisms in the 3'UTR in the neurocalcin delta gene affect mRNA stability, and confer susceptibility to diabetic nephropathy. Hum Genet 122, 397-407.

93. Shilo, L., Sabbah, H., Hadari, R., Kovatz, S., Weinberg, U., Dolev, S., Dagan, Y., and Shenkman, L. (2002). The effects of coffee consumption on sleep and melatonin secretion. Sleep Med 3, 271-273.

94. Sullivan, P.F., Fan, C., and Perou, C.M. (2006). Evaluating the comparability of gene expression in blood and brain. Am J Med Genet B Neuropsychiatr Genet 141B, 261-268.

SUPPLEMENTARY INFORMATION

Supplementary Table 1. Genotyping and Genetic Imputation information of all cohorts

Study sample	Origin	Sample type	Sample size (%women)	Genotyping Platform	Quality control of genotyped SNPs				Genetic Imputations software used	Build	Analysis software used	Total # of SNPs used for imputation	λ
					HWE p-value	SNP call rate	Sample call rate	MAF					
ERF	Dutch	Family based	1814(54)	Illumina 6K, 318K, 370K, Affymetix 250K	10^{-6}	95%	95%	0.01	MACH	36	ProbABEL	up to 487573	1.01
KORA	German	Population based	1814(51)	Affymetrix 1000K	10^{-5}	95%	95%	0.01	IMPUTE	36	SNPTEST	688765	0.99
NTR1	Dutch	Population based	732(71%)	Perlegen 600K	-	95%	95%	0.01	IMPUTE	36	SNPTEST	435291	1.01
NTR2	Dutch	Population based	355(63%)	Illumina 660K	10^{-5}	95%	95%	0.01	IMPUTE	36	SNPTEST	515781	1.01
RS-I	Dutch	Population based	4190	Illumina550K	-	90%	97.5%	-	MACH	36	ProbABEL	491875	1.02
RS-II	Dutch	Population based	1895	Illumina550K	-	90%	97.5%	-	MACH	36	ProbABEL	-	1.02
SHIP	German	Population based	2125 (77.4)	Affymetrix SNP 6.0	NA	NA	92%	NA	IMPUTE	36	QUICKTEST	869224	1.03
TwinsUKI	UK	Twins	1092 (99.73)	Illumina 317K	10^{-6}	95%	98%	0.01	IMPUTE	36	GenABEL	309197	1.00
TwinsUKII	UK	Twins	1919 (87.95)	Illumina 660K	10^{-6}	95%	98%	0.01	IMPUTE	36	GenABEL	582591	1.00
QIMR	Australian	Twins	2240 (74.1)	Illumina 317K, Illumina 370K, Illumina 610K	10^{-5}	95%	98%	0.01	MACH	36	ProbABEL	271091	1.00
LifeLines	Dutch	Population Based	7929(57)	Illumina CytoSNP12v2	10^{-5}	95%	98%	0.01	BEAGLE v3.1.0	36	PLINK	257581	1.02

Supplementary Table 2. Descriptive Statistics of all cohorts in the study

Population	sample	Average cups of coffee/day(sd)	Average age (sd)
ERF	Total	5.47(3.95)	48.62(14.49)
	Male	5.88(4.39)	49.37(14.48)
	Female	5.13(3.49)	48.04(14.48)
KORA	Total	3.34 (2.95)	53.91 (8.86)
	Male	3.44 (3.22)	54.22 (8.92)
	Female	3.26 (2.66)	53.62 (8.79)
NTR1	Total	2.79(2.96)	33.16(11.71)
	Male	4.17(3.33)	33.51(12.41)
	Female	2.23(2.59)	33.02(11.42)
NTR2	Total	2.98(3.14)	33.53(12.23)
	Male	4.33(3.71)	33.37(11.49)
	Female	2.20(2.45)	33.63(12.66)
RS-I	Total	3.85 (1.91)	70.21 (9.65)
	Male	4.16 (2.09)	68.57 (8.59)
	Female	3.63 (1.75)	71.26 (10.12)
RS-II	Total	4.36 (2.64)	65.98(10.55)
	Male	4.76(2.72)	64.74(9.49)
	Female	4.03(2.52)	66.83(11.14)
SHIP	Total	2.67 (1.95)	53.45 (15.35)
	Male	2.79 (2.10)	59.77 (12.52)
	Female	2.63 (1.91)	51.60 (15.61)
TwinsUKI	Total	1.72 (2.51)	54 (11.32)
	Male	2 (1.73)	34.97 (10.1)
	Female	1.72 (2.51)	54.05 (11.29)
TwinsUKII	Total	1.61 (2.12)	54.2 (12.68)
	Male	1.74 (2)	54.55 (13.04)
	Female	1.59 (2.14)	54.15 (12.63)
QIMR	Total	2.82(2.40)	31.46(11.14)
	Male	2.62 (2.33)	28.09 (7.94)
	Female	2.89 (2.42)	32.66 (11.55)
LifeLines	Total	4.01(2.40)	47.48(10.82)
	Male	4.67(2.47)	47.69(10.91)
	Female	3.52(2.21)	47.33(10.76)

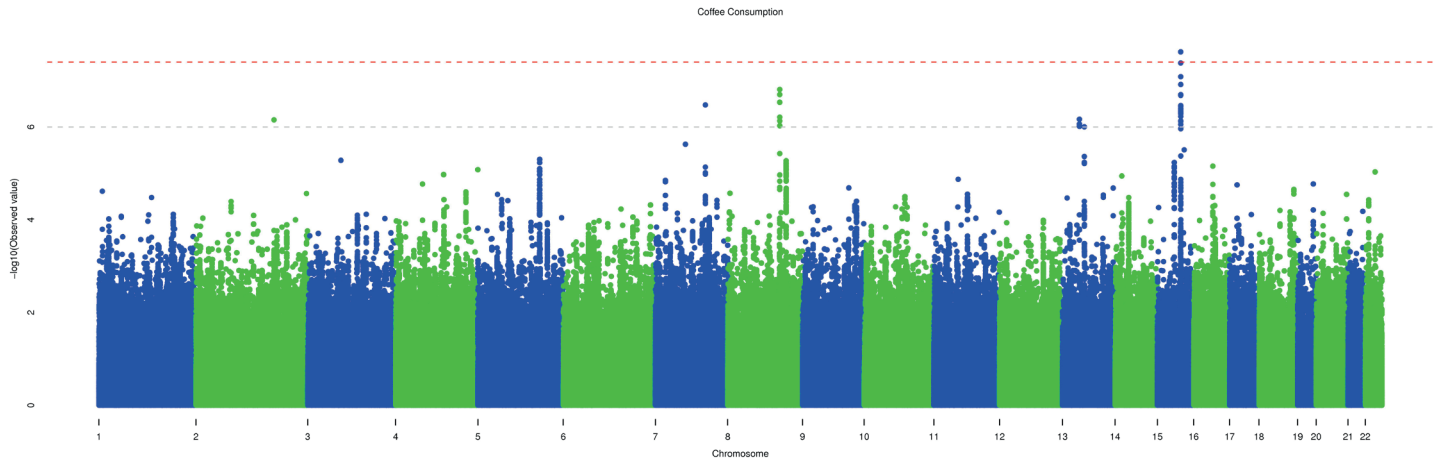
Supplementary Table 3. eQTL analysis of the top hits in GWAS

ProbeID	Pos	SNP	Allele	Effect	H2	Lod	Pvalue	Chr	gene
1557325_at	72.913	rs6495122	C	0.289	4.14	3.238	0.00011	15	---
209956_s_at	72.807	rs2470893	G	-0.335	4.62	3.299	1.00E-04	15	CAMK2B
229426_at	72.807	rs2470893	G	0.346	4.93	3.224	0.00012	15	COX5A
229426_at	72.913	rs6495122	C	-0.351	6.11	4.708	3.20E-06	15	COX5A
232016_at	72.913	rs6495122	C	0.275	3.74	3.007	2.00E-04	15	KIAA1018
240756_at	72.913	rs6495122	C	-0.282	3.94	3.197	0.00012	15	---

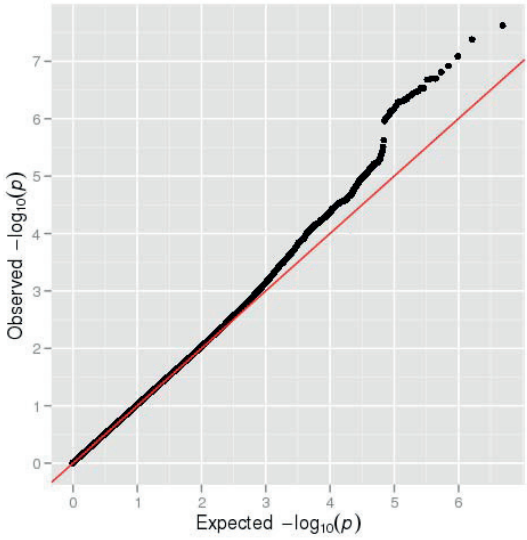
Supplementary Table 4. Additional evidence of association of the top SNPs in an independent meta-analysis of GWAS of coffee drinking among coffee drinkers performed by deCODE (unpublished data)

MarkerName	Allele1	Allele2	DGCN					NIJMEGEN				
			p-value	Beta	Se	Info	Freq	p-value	beta	se	Info	freq
rs16868941	G	A	0.80708	0.015905	0.066855	0.992196	0.751324	0.16577	0.091933	0.066743	0.989751	0.789612
rs382140	G	A	0.46269	0.048663	0.068012	1	0.778832	0.26891	0.072604	0.066077	1	0.797333
rs7600169	G	A	0.98386	0.002474	0.125543	0.832417	0.931882	0.43907	-0.07342	0.095467	0.907107	0.895951
rs9526558	G	A	0.63512	-0.03373	0.072968	0.996371	0.183748	0.55405	-0.03846	0.065401	0.996549	0.216426

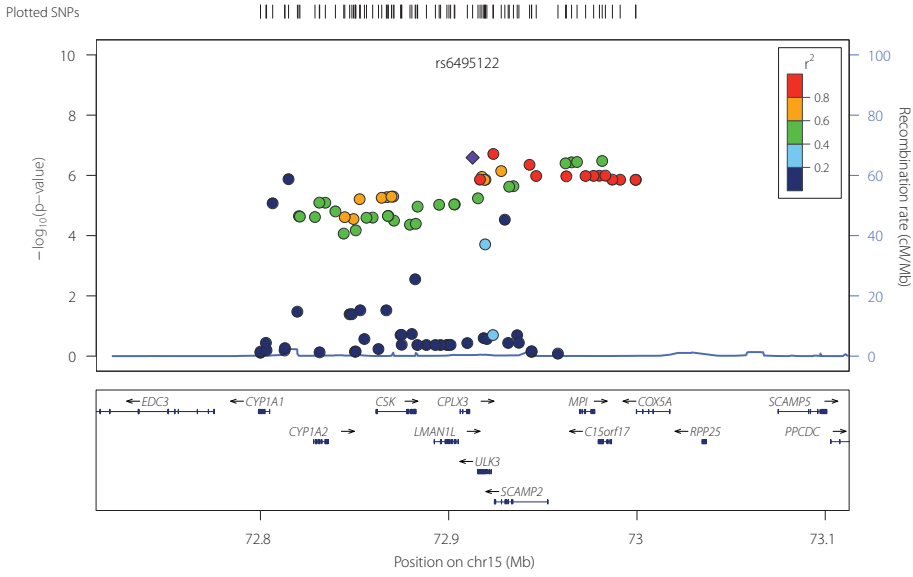
Supplementary Figure 1. Genome-wide association plot for coffee drinking. The vertical axis shows the negative logarithm of the association p-values and the horizontal axis shows the whole autosomal genome divided into 22 chromosomes. Each dot represents a SNP. The red dotted horizontal line depicts the genome-wide significance threshold p-value of $5*10^{-08}$.



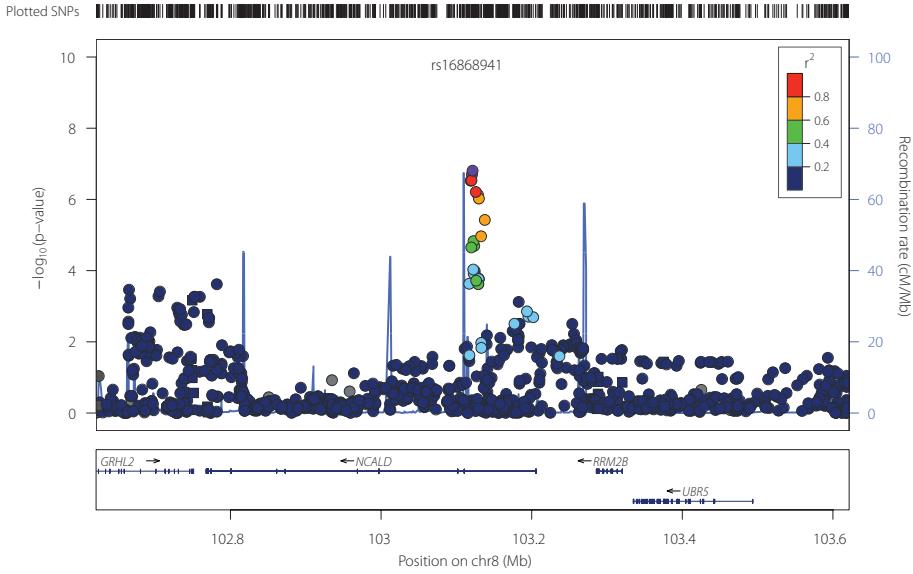
Supplementary Figure 2. Quantile-Quantile plot for Coffee drinking. The horizontal axis shows the negative logarithm of the expected p-values from a 1 d.f. chi-square distribution and the vertical axis shows the negative logarithm of the p-values from the observed chi-square distribution. Each black dot represents a SNP.



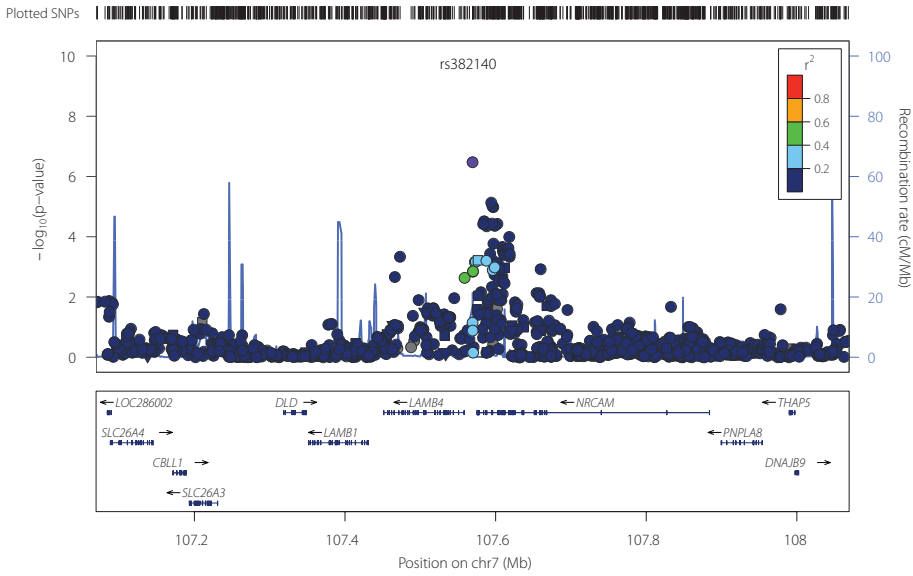
Supplementary Figure 3. Regional association plot for 15q24 in the RS-II after adjusting for the two most significant hits in the region. The vertical axis shows the negative logarithm of the association p-values and the horizontal axis shows the position in mega bases. The purple diamond represents rs6495122, circles represent other SNPs in the region with different colours representing the extent of linkage disequilibrium of these SNPs with rs6495122. Genes in the region are shown below the horizontal axis.



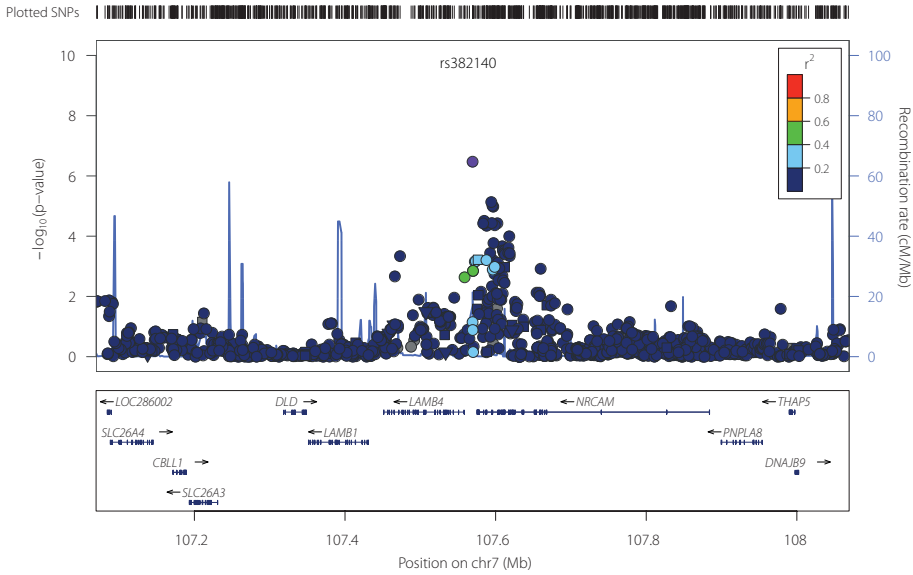
Supplementary Figure 4. Regional association plot for chromosome 8 (rs16868941). The vertical axis shows the negative logarithm of the association p-values and the horizontal axis shows the position in mega bases. Each dot represents a SNP with the purple dot representing the top SNP (rs16868941) in the region while other colours representing the extent of linkage disequilibrium of other SNPs with top SNP. Genes in the region are shown below the horizontal axis.



Supplementary Figure 5. Regional association plot for chromosome 7 (rs382140). 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 a SNP with the purple dot representing the top SNP (rs382140) in the region while other colours representing the extent of linkage disequilibrium of other SNPs with top SNP. Genes in the region are shown below the horizontal axis.



Supplementary Figure 6. Regional association plot for chromosome 13 (rs95265580). 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 a SNP with the purple dot representing the top SNP (rs95265580) in the region while other colours representing the extent of linkage disequilibrium of other SNPs with top SNP. Genes in the region are shown below the horizontal axis.



Chapter 5

General Discussion

Findings of this thesis

Behaviour refers to the actions of an organism, usually in relation to its environment. In humans, behaviour appears to be controlled primarily by the endocrine and nervous systems. The complexity of behaviour in humans is correlated to the complexity of the nervous system. Behaviours can be innate or learned. Many behaviours are highly heritable both within and across species.

In this thesis I have made an effort to unfold the genetics of traits that have a strong behavioural aspect and are known to affect social, somatic and psychological health. The traits I have studied concern a variety of outcomes including personality, ADHD, coffee drinking and sleep. All are common and/or quantitative traits and also known to be significantly heritable. I have made use of various gene mapping techniques including linkage and association. Moreover, I also use more advanced molecular and statistical analysis including copy number variation (CNV) and genomic imprinting analysis to find genetic variants associated with these traits. Using these methods I discovered novel loci (genes) implicated in the traits studied.

The first trait studied in this thesis is attention deficit hyperactivity disorder (ADHD). ADHD is a psychiatric developmental disorder[1] with an extremely strong behavioural component, which results in impaired social and/or academic functioning. In **chapter 2** we report the findings of a genome-wide linkage analysis of ADHD in nine inbred children identified in the Genetic Research in Isolated Populations (GRIP program). We found suggestive evidence of linkage (LOD = 2.2) of ADHD to a novel locus at 18q21-22 using homozygosity mapping. This finding was also supported by the genome-wide association analysis performed by the Genetic Association Information Network (GAIN) (rs2311120, p-value = $1.06*10^{-05}$). In a sub-analysis of inattentive type ADHD children we also found suggestive evidence of linkage to 6p22 (LOD=2), a region which was found earlier for ADHD in patients with reading disability[2]. Other regions with HLOD ≥ 1 include 1p36, 5q33, 6p12, 18p11, and 15q25. Although not significant, most of these regions including 1p36, 5q33 and 6p12 overlap or are adjacent to those found by the previous studies[3-5] and harbour serotonin receptors including *HTR1B, HTR1E, HTR4, HTR1D,* and *HTR6*. This suggests that linkage analysis is a powerful approach for identification of genetic loci as we were able to discover several interesting candidate regions with only nine patients compared to the 958 trios in the GAIN analysis.

I next focused on personality traits. The structure of human personality has traditionally been accounted for by a relatively small set of traits. Over the last century, scientific consensus has converged on a taxonomic model of personality traits based on five higher-order dimensions of Neuroticism, Extraversion, Openness to Experience, Agreeableness and Conscientiousness, known as the Five-Factor Model (FFM)[6]. These five dimensions are developed as orthogonal scales and therefore are largely independent. Together they provide a broad description of personality. Neuroticism is commonly defined as emotional instability; it involves the experience of negative emotions such as anxiety, depression, hostility, and

vulnerability to stress. Extraversion is characterized by positive emotions, gregariousness, and the tendency to be active, seek out stimulation and enjoy the company of others. Openness to Experience involves active imagination, aesthetic attentiveness, variety preference and intellectual curiosity. Agreeableness can be defined as the tendency to be cooperative and compassionate rather than suspicious and antagonistic towards others. Lastly, the dimension of Conscientiousness reflects self-discipline, carefulness, thoroughness, organization, deliberation and achievement. Personality traits predict a host of social, behavioural and health outcomes, such as job performance, longevity, and many psychiatric disorders, including substance abuse and dependency, mood disorders such as major depressive disorder (MDD), anxiety disorders, and personality disorders[7-20]. Although highly heritable with heritability estimates ranging between 33% and 65%[21-25], finding genes implicated in personality has proved to be difficult.

During my PhD training, I participated in a large genome-wide association study (GWA) aiming to identify common genetic variants with small effects. Included in the analyses were ten discovery samples (17,375 adults) and five in-silico replication samples (3,294 adults). All participants were of European ancestry. Personality scores for Neuroticism, Extraversion, and Openness to experience, Agreeableness, and Conscientiousness were based on the NEO Five-Factor Inventory (NEO-FFI)[6]. Genotype data were available of ~ 2.4 million single nucleotide polymorphisms (SNPs; both directly typed and imputed using HAPMAP data). In the discovery samples, classical association analyses were performed under an additive model followed by meta-analysis using the weighted inverse variance method. Results showed genome-wide significance for Openness to experience near the *RASA1* gene on 5q14.3 (rs1477268 and rs2032794, p-value = $2.8*10^{-8}$ and p-value = $3.1*10^{-8}$) and for Conscientiousness in the brain-expressed *KATNAL2* gene on 18q21.1 (rs2576037, p-value = $4.9*10^{-8}$). We further conducted a gene-based test that confirmed the association of *KATNAL2* to Conscientiousness. In-silico replication did not, however, show significant associations of the top SNPs with Openness and Conscientiousness, although the direction of effect of the *KATNAL2* SNP on Conscientiousness was consistent in all replication samples. Larger scale GWA studies and alternative approaches are required for confirmation of *KATNAL2* as a novel gene affecting Conscientiousness.

For this thesis I followed an alternative approach. To find genetic loci that may confer moderate to large effects on personality traits of Neuroticism, Extraversion, Openness, Agreeableness and Conscientiousness, in **chapter 3.1** I performed qualitative genome-wide linkage scans of high scoring individuals (> 90^{th} percentile) for each of the five traits in the Erasmus Rucphen Family (ERF) Study. We found significant evidence of linkage of Conscientiousness to 20p13 under the recessive model (log of odds (LOD) = 5.86). Haplotype construction revealed a distinct haplotype segregating in six high LOD score contributing families, which showed marginal association (p-value = 0.058 and 0.038) under a recessive model. This haplotype covers a 100 kilobase (kb) region on 20p13 and contains 5 genes belonging to the beta defensin family (*DEFB*). We also found suggestive evidence of linkage of Neuroticism to 21q22 (LOD = 3.42),

where we found two significant haplotypes (p-values = 0.009, 0.007), one of which was being shared in a homozygous state by most of the high scoring individuals from three high LOD score contributing families. This region harbors 10 genes, including *TTC3, DSCR9, DSCR3, DYRK1A, KCNJ6* and *PIGP*, and has been associated with cognitive and mental impairment in individuals with Down's syndrome[26]. Suggestive linkage (LOD = 3.32) yet a significant haplotype (p-value = 0.018) was found for Agreeableness in the 17q24 region. Genes in this region include *KCNJ2, KCNJ16* and *MAP2K6*. Excess homozygosity in each of these linked regions prompted us to look for deletions in the regions. However, the copy number variation (CNV) analysis did not provide evidence for CNVs. These regions did not emerge in the GWAS I mentioned earlier. There may be several explanations for this including a false positive finding in this study. However, more likely, GWA studies are unlikely to pick up relatively rare recessive variants[27]. Such variants may indeed explain part of the missing heritability in quantitative traits[27].

In **chapter 3.2** we performed an alternative linkage analysis including the various personality outcomes quantitatively. Although quantitative traits are a powerful approach, the number of subjects to be studied is critical. Further, also for linkage analysis replication is important to exclude false positive findings. Therefore, a meta-analysis of four un-published quantitative trait linkage analyses of the five personality traits was conducted. The samples (*N* = 6,149) included Erasmus Rucphen Family Study (ERF), Netherlands Twin Register (NTR) and an adult and adolescent sample from Australia (QIMR). To maximize power we combined genome-wide linkage and association analyses methods in a two step analysis [28]. The regions of significant and suggestive evidence of linkage were fine mapped using the results of the largest meta-analysis of genome-wide association studies (*N* ~ 18,000) of these five personality traits. Significant linkage was detected for Neuroticism at chromosome 3p14 (LOD = 4.67) and at chromosome 19q13 (LOD = 3.55), for Extraversion at 14q32 (LOD = 3.3); for Agreeableness at 3p25 (LOD = 3.67) and at two adjacent regions on chromosome 15 including 15q13 (LOD = 4.07) and 15q14 (LOD = 3.52) in the individual scans. In the meta-analysis, we found strong evidence of linkage of Extraversion to 4q34, 9q34, 10q24 and 11q22, Openness to 2p25, 3q26, 9p21, 11q24, 15q26 and 19q13, and Agreeableness to 4q34 and 19p13. When combining these data with the association results of the GWAS of these personality traits, significant evidence of association was detected for Openness at 11q24 (rs677035, p-value = $2.6*10^{-06}$). Borderline evidence for association was detected between Neuroticism and rs332389 (p-value = $9*10^{-05}$) at 3p14 and between Extraversion and rs7088779 (p-value = $4.2*10^{-06}$) at 10q24. Of these three SNPs, rs677035 is an intergenic SNP located between *FLI1* and *KCNJ1*, rs332389 is intronic to the gene *SLC25A26 a*nd rs7088779 is located between *CRTAC1* (cartilage acidic protein 1) and *C10orf28*; a region previously implicated in Alzheimer's disease. Also in the quantitative linkage analysis, which is more similar to GWAS, there was no evidence for a role of *KATNAL2* or *RASA1* for Conscientiousness and Openness. Again one may argue that this may be explained by a false positive finding or the differences in power for the various variants implicated in these

complex traits. An interesting gene family that emerges from the linkage studies in this thesis is the KCNJ family. The linkage regions we identified using both dichotomous and quantitative outcome are enriched with KCNJ genes including *KCNJ6, KCNJ16, KCNJ2* and *KCNJ1*. KCNJs are inwardly rectifying potassium channels. They are the targets of multiple toxins, and malfunction of the channels has been implicated in several diseases including neurological and endocrine disorders[29] and electrocardiogram arrhythmias[30].

Epigenetic effects, of which genomic imprinting is a form, is often cited as a reason for the loss of efficiency of genome-wide association studies in identifying common variants for complex traits[31]. In **chapter 3.3** we performed a genome-wide search for imprinted regions for personality traits in the ERF sample using SNPs from the Illumina 6K linkage panel. We found significant differences in paternally and maternally transmitted alleles at 6q25 (rs602890, p-value = $5*10^{-07}$) for Extraversion and at 4q21 (rs1880719, p-value = $8*10^{-07}$) for Agreeableness. The two SNPs, which did not show any significance in a general association test, showed significant association when parental origin of alleles was incorporated in the association model (p-value= $3*10^{-06}$ for rs602890 and p-value = $1*10^{-06}$ for rs1880710). Rs602890 showed opposite and nominally significant effects when inherited from father (β = -2.78, p-value = $2.7*10^{-03}$) and mother (β = 4.33; p-value = $2*10^{-04}$) and rs1880719 showed significant association with Agreeableness only when inherited from the mother (β = -4.47, p-value = $6*10^{-05}$). The 6q25 region is a known imprinted region[32], while imprinting is also hypothesized as the reason for the central nervous system overgrowth in the 4q21/4q23 syndrome[33]. Rs602890 is an intronic SNP in the gene *ZDHHC14* and rs1880719 is an intronic SNP in the gene *SCD5* encoding Stearoyl-CoA desaturase. None of these SNPs showed any evidence of association in the large GWAS or linkage studies when ignoring parent-of-origin effects. These findings can only be replicated in a family based sample. An attempt to replicate the two most interesting findings (rs1880719 and rs602890) in the Australian twin samples failed to provide sufficient evidence of parent-of-origin effects. We assume that this may be due to the sparse set of markers that we used in our analysis (5,250), making it difficult to pinpoint the region of interest precisely. A fine mapping of the 4q21 region revealed four additional significant SNPs (in high LD with rs1880719) in the same gene (*SCD5*). At present another replication effort is being made in the twin sample from the Netherlands Twin Registry (NTR).

Finally, we have studied sleep and caffeine in the thesis. Epidemiological studies have shown that sleep duration is associated with psycho-somatic malfunctioning[34-37]. In **chapter 4.2** we performed the first large scale genome-wide association analysis of sleep duration in 7 European cohorts (N = 4,260). We identified an intronic variant on chromosome 12p12.1 (rs11046205) in the *ABCC9* gene and validated it in a replication study. When the conserved *Drosophila* homologue was knocked down pan-neuronally by RNAi, night sleep duration was reduced. *ABCC9* encodes an ATP-sensitive potassium channel subunit (*SUR2*), which functions as a sensor of intracellular energy. Our findings therefore link sleep and its restorative effects,

known in physiological terms as the homeostat, to cellular energy metabolism. Time to fall asleep (sleep latency) is a major determinant of the quality of sleep. Increased sleep latency is one of the major characteristic of sleep onset insomnia[38] and delayed sleep phase syndrome[39], which are two common circadian disorders that hamper an individual's social and academic life and work performance. With no previous genetic studies performed on this trait, in **chapter 4.1** we performed a meta-analysis of seven genome-wide association studies consisting of 4,270 individuals of Caucasian origin. We found strong evidence of association of a variant (rs9900428, p-value = $5.7*10^{-08}$) of the hexaribonucleotide binding protein 3 (*HRNBP3*) (also known as *FOX3*) with sleep latency. We successfully replicated our findings in five additional cohorts ($N = 10,901$). The meta-analysis of discovery and replication cohorts showed a genome-wide significant association of rs9900428 (p-value = $3.8*10^{-08}$). The *FOX3* gene is a highly brain specific protein that belongs to the family of FOX genes, which are regulators of pre and postnatal alternative splicing in muscle and brain tissues. *FOX3* is suggested to be an upstream regulator for alternative splicing of brain specific mRNAs involved in glutamatergic and GABAergic transmission. This is the first genome search on sleep latency and no variant for sleep latency has been reported earlier.

We finally study a major determinant of sleep latency, caffeine. Caffeine is known to have behavioural effects when taken in moderate amounts[40-42]. High doses of caffeine can have also negative effects such as anxiety, restlessness and insomnia. Caffeine intake is also described as a model of drug abuse[43] and coffee is the major source of caffeine in adults. In **chapter 4.3** we performed a large scale meta-analysis of genome-wide association studies on coffee intake. Our data ($N \sim 18,000$) consisted of samples from eight cohorts. We found genome-wide significant evidence of association of a variant (rs2470893) in 15q25 region, intergenic to genes *CYP1A1* and *CYP1A2*. The "T" allele of rs2470893 was positively associated with coffee drinking. The *CYP1A1* gene was also found to be down-regulated after caffeine treatment in lymphoblastoid cell lines. *CYP1A1* is known to metabolize polycyclic aromatic hydrocarbons such as benzo(a)pyrene, which is an important constituent of coffee and also known to be involved in the mutagenesis resulting from coffee intake while *CYP1A2* is involved in primary metabolism of caffeine. Combined evidence from GWAS and caffeine expression analysis suggests that *CYP1A1* regulates coffee drinking and that this association is independent of smoking status. Apart from *CYP1A1/CYP1A2* we also found strong evidence of association of coffee intake with polymorphisms in *NCALD* (rs16868941), *NRCAM* (rs382140) and *CAB39L* (rs9526558) genes. Earlier we found strong association of the *NCALD* gene with sleep latency (rs17498920, p-value = $2*10^{-05}$). This is an interesting finding since coffee intake is known to interfere with melatonin secretion and delay the onset of sleep. A replication analysis of the top findings from the GWAS in a further 8,000 individuals of Dutch origin (LifeLines), confirmed association of *CYP1A1/A2* locus and rs382140 near the *NRCAM* gene with coffee drinking. *CAB39L*, although not successfully replicated was found to be up-regulated in all the three cell lines studied after caffeine treatment suggesting

that the gene is indeed involved in caffeine metabolism and intake. Association analysis of the polymorphisms in these genes with coffee related phenotypes including blood pressure, mortality, cancer mortality, lipids and Alzheimer disease (AD) revealed significant association of rs2470893 (*CYP1A1/1A2*) with systolic and diastolic blood pressure, where again "T" allele was positively associated, suggesting that increased coffee intake elevates blood pressure[44]. The association of our top hits with blood pressure is very interesting and strengthens the argument that coffee consumption elevates blood pressure[45].

In search of the missing heritability

For the past 5 years genome-wide association studies (GWAS) have dominated the search for new genes for complex diseases overtaking other approaches of gene finding such as candidate gene and linkage analyses. Facilitated by technological developments in molecular biology, genetic epidemiologists have so far discovered many variants associated with several common diseases and traits such as Type 2 Diabetes, age-related macular degeneration and Crohn's disease[46]. There currently are 26 established susceptibility genes published for type 2 diabetes[47] 54 for human height and 22 for lipid levels[48-49]. These variants still explain only a small part of the genetic variance or heritability, for human height and lipids up to 4–6%[50-51], and subsequently the search for novel variants continues to unravel 'missing heritability'.

This missing heritability is explained by additional rare variants with strong effects and/or common variants with weak effects, acting additively and/or interacting with other genetic and environmental variants. To discover these additional genetic factors, GWAS need to enlarge, and this has led to further expansion of existing consortia and the establishment of new ones. Since the first publication in 2005[52], GWAS have undergone enormous evolution: from 10,000 single nucleotide polymorphisms (SNPs) in 100 individuals of a single sample[52] to 1 million genotyped and ~ 2.5 million imputed SNPs in more than 80,000 individuals of multiple samples [53]. The decreasing costs of genotyping, new statistical methodologies, and increasing willingness of the scientists to share and pool data sets have facilitated these rapid developments and made this approach very successful also in the setting of epidemiology. For instance, the Cohort for Health and Aging Research (CHARGE) is studying multiple common traits in 50,000–70,000 individuals from US and European follow-up studies[54-56], and the Dutch three-generation study LifeLines is going to include 165,000 participants[57]

While increasing size will help in finding new variants with smaller effects, there will also be true positives that remain undetected in the larger consortia because of the stringent threshold levels of statistical significance imposed in GWAS (p-value < 5 *10^{-08}) to adjust for multiple testing. The chances of success of consortia are further reduced if confounding due to population heterogeneity, also refer to as population admixture, is to be adjusted for, which is the case when populations are of different genetic origins. Therefore new approaches are needed to identify genetic variants explaining the missing heritability and one such new approach was

used successfully in a recent GWAS in schizophrenia that was published in *Nature*[58]. The classical GWAS analysis produced only one genome-wide significant polymorphism, but the authors used a new 'genetic scoring' method through which they demonstrated that there indeed existed undetected variants below the threshold. How to detect variants that are not detected? Basically, the method tests the association of a score variable that manifests a combined effect of many SNPs. The polymorphisms in the score are selected on the basis of their nominal p-value in the predefined discovery sample. Scores can be generated for any arbitrarily chosen threshold of nominal statistical significance, for instance selecting all SNPs with p-values lower than e.g. 0.01, 0.1 or 0.5. The significance of the score is then tested by using it as a predictor in a simple regression model in an independent 'target sample'. In this target sample, a one-parameter test for all SNPs can be used, thus relaxing the conservative p-value of $5*10^{-08}$ needed for testing all SNPs in GWAS to classical significance level of 0.05. Using data from the International Schizophrenia Consortium with men defined as the discovery sample and women as the target sample, the authors showed that a score based on all SNPs with p-value < 0.5 was most strongly and significantly correlated with schizophrenia in the target sample compared to the scores based on other thresholds. The fact that the set of SNPs with p-value < 0.5, including both many falsely and an unknown number of truly associated SNPs, predicted better than the score with p-value < $5*10^{-08}$ suggests that both the number of undetected relevant variants as well as their joint effect on the outcome is substantial[58]. The authors further showed that the score correlated significantly with related diseases as bipolar disorder, but not with unrelated outcomes such as Crohn's disease, coronary artery disease, hypertension, rheumatoid arthritis or type 1 and type 2 diabetes. This suggests that schizophrenia and bipolar disorder have a shared genetic component and also that the selected alleles were specific to schizophrenia and related disorders[58].

The genetic scoring method is logical and simple as among the SNPs that fail to reach the significance threshold in the GWAS there ought to be true associations, which just do not reach the threshold because the study does not have enough power[59]. There may, however, be several caveats. First, the informative value of the approach depends on the size of the discovery sample. If the discovery sample is small, more falsely associated SNPs will be selected at each threshold, and consequently scores do not explain much of the phenotypic variance in the target sample. The second caveat is that also a score based on 38,000 SNPs with a p-value lower than 0.5, derived in a discovery sample of 3,800 individuals, explained only 3% of the heritable variance in the target population of 3,100 persons. It can be expected that a larger discovery set will select more true positives among those with a p-value lower than 0.5 and therefore explain a higher percentage of the variance in the target sample. However, simulations showed that the variance explained by the scores can increase from 3 to 20% if the size of the discovery sample is increased to 20,000 individuals[58]. Thus, also for this new method the size of the discovery sample is an important determinant of success. Third, one of the major conclusions on the basis of this

method is that there are undetected common genetic contributions. Of course one may argue that this observation could already be inferred from the fact that there is 'missing heritability'. But perhaps an even more important limitation of the genetic scoring method is that it does not tell which one(s) of the variants included is responsible for the statistical significance.

Then what can we do with this information? First, the method may be used to improve our understanding the genetic architecture of the disease or trait. Scores can be calculated and tested for multiple different significance thresholds levels of statistical significance. By comparing the proportions of explained variance across these thresholds, a pattern may be observed. When going up from a very low threshold, e.g., p-value < 10^{-07} to p-value < 0.5, we may see that scores may rise to a certain point and then either decline or become stable, a pattern which suggests that a few genes with stronger effects may be involved. When the proportion of explained variance monotonically increases until all SNPs are included in the scores, there are likely to be a large number of common variants with small effects. So the scores calculated over several different cut offs can give an indication on how complex the trait is, on the likelihood that the trait has a polygenic basis. For example, for schizophrenia the score goes up from 0.004 to 0.025 by moving up from a threshold of p-value < 0.01 to p-value < 0.5[58], which is an indication that many more common low risk variants are likely involved in schizophrenia.

Second, this method could be considered as an intermediate step in the gene discovery process. When scores are statistically significant, one may consider to only analyzing the included SNPs in the independent samples. For replication purposes this leads to a less stringent level of statistical significance, and potentially to a higher likelihood of finding susceptibility variants. Because the success of this approach will depend on the size of the discovery sample — the larger the discovery sample the more likely true susceptibility genes will be selected in the scores — its added value of selecting SNPs in much smaller independent populations may not be efficient. More promising is to use the score approach to select SNPs for use in complex modelling of the trait for instance to study gene by gene interactions which otherwise seems impossible with 2.5 million SNPs.

Third, the method could be used to predict disease for preventive and clinical purposes. Evans and colleagues applied the score approach and assessed the discriminative ability for several threshold levels of statistical significance in several complex diseases [60]. When significance thresholds were varied from 10^{-05} to 0.8, discriminative ability improved for bipolar depression, coronary heart disease, hypertension and type 2 diabetes, but decreased for rheumatoid arthritis and type 1 diabetes prediction. For all diseases, the discriminative ability was lower than what would be obtained when testing known susceptibility genes, except for hypertension where no susceptibility variants were known at the time and for bipolar disorder for which the score performed better than the known variants, but only for the liberal and not for the stringent significance thresholds. It is also suggested that the shared genetic liability between schizophrenia and bipolar disorder would make the genetic based refinement of the

diagnosis of these diseases possible, which may also be tried for other diseases with overlapping symptoms[58]. Of all potential applications of the genetic scoring method, this is the least substantiated, and it may still be too premature[61]. However, when the proportion of variance explained by the scores can increase from 3 to 20%, as suggested by simulation analyses by improving the power of the discovery set[58], the discriminative accuracy could be in the range of what we commonly see for non-genetic risk prediction models in e.g., cardiovascular diseases, diabetes and mortality[62-64].

GWAS have been very successful in finding multiple variants for many traits, but we are reaching the limits of what can be found through this approach sooner or later. Whether the genetic scoring method will be successful in finding more risk variants for complex traits and in unravelling 'missing heritability' remains to be determined. So far we applied this method successfully in depression[65], lipids and glaucoma[66]. Although the approach works, the variance explained is still very low, limiting the value for application in public health or decision making. More recently a new approach has been proposed which uses the restricted maximum likelihood modelling (REML) to estimate the variance explained by the SNPs[67]. The authors show that using REML and kinships estimated from the genotyped SNPs, they could explain roughly half of the heritability of human height using common SNPs. For the remaining half they suggest incomplete LD between causal variant and the SNPs as the major limitation, which, they suggest, might be overcome using larger discovery samples. An issue to be evaluated is the extent to which the data over fits the regression model, limiting the predictive value to the population studied. It is clear that such over fitting makes the approach useless in the field of epidemiology or public health. The new genetic score method is one approach, approaches aiming at testing of more complex models with gene by gene and gene by environment interactions may be another avenue, although it is clear that epistasis and gene-environment interactions do not form a part of the missing heritability. Last but not least technological developments may come to rescue with new development in high throughput sequencing.

Future research

With whole-genome sequencing of hundreds of individuals within our reach, contemporary thinking in genetic epidemiology is distributed within two extremes: continue enlarging the GWAS to find more variants using the SNPs or to go forward with whole genome/exome sequencing with a view to find rare variants. The proponents of the former suggest that given enough sample size the GWAS will be able to find most of the genetic variants[67] associated with a certain trait or in other words people who strictly adhere to the common disease/common variant theory. Others suggest that we have reached the limit with the GWAS[68-69] and profess sequencing. As expected and based upon population genetics, my thesis shows that the two lines of thought are not mutually exclusive and should converge with infinitively large samples. We have witnessed overlap of our linkage findings for ADHD and personality traits with those of the respective GWAS.

If larger GWAS are to be performed, more complex modelling including parent-of-origin effects may be useful as we and others have shown[70]. Currently only family based studies are capable of performing such analysis. A limitation of these studies is that the replication samples are usually not available. However, more recently developed long range phasing method[71] may enable population based studies to incorporate parent-of-origin effects in the association models. Also for the complex traits, reducing the complexity of the trait, that is, the use of simpler base level phenotypes is another approach which can improve the performance/power of the GWAS to detect common variants. This is already being done for a variety of complex traits/disorders. For instance, each of the five personality traits mentioned earlier in this thesis are defined by six underlying sub-phenotypes. We are now performing the GWAS on each of the individual sub-phenotypes in an ongoing study. Expansion of the GWAS in terms of the number of SNPs used is yet another avenue which might lead to the discovery of more of the common variation. Most consortia are now switching to 1000 genomes[72] as the reference set for SNP imputations rather than the HapMap. This will increase the number of SNPs from 2.5 to 8 million to be used in the GWAS, thus improving the chances of finding more common variants.

Needless to say that whole genome/exome sequencing will reveal the rare variants in the genome, however, the enthusiasm is also coupled with a feeling of anxiety about what sequencing will reveal and whether or not we are capable of handling such an enormous amount of information with contemporary bioinformatics and methodological tools. Provided the technological and methodological advances, the ultimate goal is to do a whole genome sequencing of all the enrolled subjects and to perform genome-wide association studies of the 3 billion base pairs[69], meanwhile designs that are optimized to detect rare variants in smaller samples are suggested[69]. These include sequencing affected individuals in families and extreme-trait sequencing. Targeted sequencing of interesting regions might be something to start with. So far linkage studies have discovered several regions for many psychiatric and non-psychiatric traits. These regions are potentially interesting targets for sequencing and might reveal rare

mutations[73]. Homozygosity mapping may also be used as a tool to provide interesting regions for targeted sequencing[74].

The findings in this thesis especially the linkage findings are interesting targets for deep sequencing. Apart from these there are a number of findings in this thesis that require a follow up. First, is the KCNJ gene family that is linked to personality traits. These genes should be characterized in more detail in humans e.g. by sequencing. These genes may also link personality to related traits, for instance, heart and electrocardiogram (ECG) traits. For Sleep, the genes that we have identified (*ABCC9* for sleep duration and *HRNBP3*/*FOX3*) also require further investigation. Although, we successfully authenticate our findings by either performing gene knockout study or through replicating in other populations, the biological mechanisms through which these genes influence sleep traits remain elusive and need to be explored. For the genes that we have discovered for coffee intake, it will be of interest to link the findings to coffee related traits and diseases. Another interesting avenue will be to look at these genes from the perspective of addiction.

REFERENCES

1. Zwi, M., Ramchandani, P., and Joughin, C. (2000). Evidence and belief in ADHD. BMJ 321, 975-976.
2. Willcutt, E.G., Pennington, B.F., Smith, S.D., Cardon, L.R., Gayan, J., Knopik, V.S., Olson, R.K., and DeFries, J.C. (2002). Quantitative trait locus for reading disability on chromosome 6p is pleiotropic for attention-deficit/hyperactivity disorder. Am J Med Genet 114, 260-268.
3. Arcos-Burgos, M., Castellanos, F.X., Pineda, D., Lopera, F., Palacio, J.D., Palacio, L.G., Rapoport, J.L., Berg, K., Bailey-Wilson, J.E., and Muenke, M. (2004). Attention-deficit/hyperactivity disorder in a population isolate: linkage to loci at 4q13.2, 5q33.3, 11q22, and 17p11. Am J Hum Genet 75, 998-1014.
4. Zhou, K., Asherson, P., Sham, P., Franke, B., Anney, R.J., Buitelaar, J., Ebstein, R., Gill, M., Brookes, K., Buschgens, C., et al. (2008). Linkage to chromosome 1p36 for attention-deficit/hyperactivity disorder traits in school and home settings. Biol Psychiatry 64, 571-576.
5. Ogdie, M.N., Fisher, S.E., Yang, M., Ishii, J., Francks, C., Loo, S.K., Cantor, R.M., McCracken, J.T., McGough, J.J., Smalley, S.L., et al. (2004). Attention deficit hyperactivity disorder: fine mapping supports linkage to 5p13, 6q12, 16p13, and 17p11. Am J Hum Genet 75, 661-668.
6. Costa, P.T., and McCrae, R.R. (1992). Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI) Professional Manual. Psychological Assessment Resources.
7. Bienvenu, O.J., Hettema, J.M., Neale, M.C., Prescott, C.A., and Kendler, K.S. (2007). Low extraversion and high neuroticism as indices of genetic and environmental risk for social phobia, agoraphobia, and animal phobia. Am J Psychiatry 164, 1714-1721.
8. Budaev, S.V. (1999). Personality and Individual Differences. Elsevier 26, 801-813.
9. Cramer, D. (1991). Type A behaviour pattern, extraversion, neuroticism and psychological distress. Br J Med Psychol 64 (Pt 1), 73-83.
10. Crowe, M., Andel, R., Pedersen, N.L., Fratiglioni, L., and Gatz, M. (2006). Personality and risk of cognitive impairment 25 years later. Psychol Aging 21, 573-580.
11. Distel, M.A., Trull, T.J., Willemsen, G., Vink, J.M., Derom, C.A., Lynskey, M., Martin, N.G., and Boomsma, D.I. (2009). The five-factor model of personality and borderline personality disorder: a genetic analysis of comorbidity. Biol Psychiatry 66, 1131-1138.
12. Goodwin, R.D., Castro, M., and Kovacs, M. (2006). Major depression and allergy: does neuroticism explain the relationship? Psychosom Med 68, 94-98.
13. Kendler, K.S., Gatz, M., Gardner, C.O., and Pedersen, N.L. (2006). Personality and major depression: a Swedish longitudinal, population-based twin study. Arch Gen Psychiatry 63, 1113-1120.
14. Meins, W., and Dammast, J. (2000). Do personality traits predict the occurrence of Alzheimer's disease? Int J Geriatr Psychiatry 15, 120-124.
15. Menza, M. (2000). The personality associated with Parkinson's disease. Curr Psychiatry Rep 2, 421-426.
16. Robertson, D.A., Ray, J., Diamond, I., and Edwards, J.G. (1989). Personality profile and affective state of patients with inflammatory bowel disease. Gut 30, 623-626.
17. Sutin, A.R., Scuteri, A., Lakatta, E.G., Tarasov, K.V., Ferrucci, L., Costa, P.T., Jr., Schlessinger, D., Uda, M., and Terracciano, A. (2010). Trait antagonism and the progression of arterial thickening: women with antagonistic traits have similar carotid arterial thickness as men. Hypertension 56, 617-622.
18. Sutin, A.R., Terracciano, A., Deiana, B., Naitza, S., Ferrucci, L., Uda, M., Schlessinger, D., and Costa, P.T., Jr. (2010). High neuroticism and low conscientiousness are associated with interleukin-6. Psychol Med 40, 1485-1493.
19. Sutin, A.R., Terracciano, A., Deiana, B., Uda, M., Schlessinger, D., Lakatta, E.G., and Costa, P.T., Jr. (2010). Cholesterol, triglycerides, and the Five-Factor Model of personality. Biol Psychol 84, 186-191.
20. Terracciano, A., Lockenhoff, C.E., Crum, R.M., Bienvenu, O.J., and Costa, P.T., Jr. (2008). Five-Factor Model personality profiles of drug users. BMC Psychiatry 8, 22.
21. Jang, K.L., Livesley, W.J., and Vernon, P.A. (1996). Heritability of the big five personality dimensions and their facets: a twin study. J Pers 64, 577-591.

22. Jang, K.L., McCrae, R.R., Angleitner, A., Riemann, R., and Livesley, W.J. (1998). Heritability of facet-level traits in a cross-cultural twin sample: support for a hierarchical model of personality. J Pers Soc Psychol 74, 1556-1565.

23. Loehlin, J.C., McCrae, R.R., Paul T. Costa, j., and John, O.P. (1998). Heritabilities of common and measure-specific components of the big five personality factors. Journal of research in personality 32, 431-453.

24. Riemann, R., Angleitner, A., and Strelau, J. (1997). Genetic and environmental influences on personality: A study of twins reared together using the self- and peer report NEO-FFI scales. J Personal 65, 449-475.

25. Van Os, J., and Jones, P.B. (2001). Neuroticism as a risk factor for schizophrenia. Psychol Med 31, 1129-1134.

26. Altafaj, X., Dierssen, M., Baamonde, C., Marti, E., Visa, J., Guimera, J., Oset, M., Gonzalez, J.R., Florez, J., Fillat, C., et al. (2001). Neurodevelopmental delay, motor abnormalities and cognitive deficits in transgenic mice overexpressing Dyrk1A (minibrain), a murine model of Down's syndrome. Hum Mol Genet 10, 1915-1923.

27. Maher, B. (2008). Personal genomes: The case of the missing heritability. Nature 456, 18-21.

28. Sham, P.C., Cherny, S.S., Purcell, S., and Hewitt, J.K. (2000). Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. Am J Hum Genet 66, 1616-1630.

29. Abraham, M.R., Jahangir, A., Alekseev, A.E., and Terzic, A. (1999). Channelopathies of inwardly rectifying potassium channels. FASEB J 13, 1901-1910.

30. Priori, S.G., Pandit, S.V., Rivolta, I., Berenfeld, O., Ronchetti, E., Dhamoon, A., Napolitano, C., Anumonwo, J., di Barletta, M.R., Gudapakkam, S., et al. (2005). A novel form of short QT syndrome (SQT3) is caused by a mutation in the KCNJ2 gene. Circ Res 96, 800-807.

31. Eichler, E.E., Flint, J., Gibson, G., Kong, A., Leal, S.M., Moore, J.H., and Nadeau, J.H. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet 11, 446-450.

32. Paterson, A.D., Naimark, D.M., and Petronis, A. (1999). The analysis of parental origin of alleles may detect susceptibility loci for complex disorders. Hum Hered 49, 197-204.

33. Nowaczyk, M.J., Teshima, I.E., Siegel-Bartelt, J., and Clarke, J.T. (1997). Deletion 4q21/4q22 syndrome: two patients with de novo 4q21.3q23 and 4q13.2q23 deletions. Am J Med Genet 69, 400-405.

34. Taheri, S., and Thomas, G.N. (2008). Is sleep duration associated with obesity-where do U stand? Sleep Med Rev 12, 299-302.

35. Stickgold, R. (2001). Toward a cognitive neuroscience of sleep. Sleep Med Rev 5, 417-421.

36. Gangwisch, J.E. (2009). Epidemiological evidence for the links between sleep, circadian rhythms and metabolism. Obes Rev 10 Suppl 2, 37-45.

37. Ikehara, S., Iso, H., Date, C., Kikuchi, S., Watanabe, Y., Wada, Y., Inaba, Y., Tamakoshi, A., and Group, J.S. (2009). Association of sleep duration with mortality from cardiovascular disease and other causes for Japanese men and women: the JACC study. Sleep 32, 295-301.

38. Eddy, M., and Walbroehl, G.S. (1999). Insomnia. Am Fam Physician 59, 1911-1916, 1918.

39. Weitzman, E.D., Czeisler, C.A., Coleman, R.M., Spielman, A.J., Zimmerman, J.C., Dement, W., Richardson, G., and Pollak, C.P. (1981). Delayed sleep phase syndrome. A chronobiological disorder with sleep-onset insomnia. Arch Gen Psychiatry 38, 737-746.

40. Mitchell, P.J., and Redman, J.R. (1992). Effects of caffeine, time of day and user history on study-related performance. Psychopharmacology (Berl) 109, 121-126.

41. Silverman, K., Mumford, G.K., and Griffiths, R.R. (1994). Enhancing caffeine reinforcement by behavioral requirements following drug ingestion. Psychopharmacology (Berl) 114, 424-432.

42. Horne, J.A., and Reyner, L.A. (1996). Counteracting driver sleepiness: effects of napping, caffeine, and placebo. Psychophysiology 33, 306-309.

43. Strain, E.C., Mumford, G.K., Silverman, K., and Griffiths, R.R. (1994). Caffeine dependence syndrome. Evidence from case histories and experimental evaluations. JAMA 272, 1043-1048.

44. Palatini, P., Ceolotto, G., Ragazzo, F., Dorigatti, F., Saladini, F., Papparella, I., Mos, L., Zanata, G., and Santonastaso, M. (2009). CYP1A2 genotype modifies the association between coffee intake and the risk of hypertension. J Hypertens 27, 1594-1601.

45. Palatini, P., Dorigatti, F., Santonastaso, M., Cozzio, S., Biasion, T., Garavelli, G., Pessina, A.C., and Mos, L. (2007). Association between coffee consumption and risk of hypertension. Ann Med 39, 545-553.

46. Manolio, T.A., Brooks, L.D., and Collins, F.S. (2008). A HapMap harvest of insights into the genetics of common disease. J Clin Invest 118, 1590-1605.

47. Zeggini, E., Scott, L.J., Saxena, R., Voight, B.F., Marchini, J.L., Hu, T., de Bakker, P.I., Abecasis, G.R., Almgren, P., Andersen, G., et al. (2008). Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet 40, 638-645.

48. Aulchenko, Y.S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I.M., Pramstaller, P.P., Penninx, B.W., Janssens, A.C., Wilson, J.F., Spector, T., et al. (2009). Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat Genet 41, 47-55.

49. Aulchenko, Y.S., Struchalin, M.V., Belonogova, N.M., Axenovich, T.I., Weedon, M.N., Hofman, A., Uitterlinden, A.G., Kayser, M., Oostra, B.A., van Duijn, C.M., et al. (2009). Predicting human height by Victorian and genomic methods. Eur J Hum Genet 17, 1070-1075.

50. Isaacs, A., Sayed-Tabatabaei, F.A., Aulchenko, Y.S., Zillikens, M.C., Sijbrands, E.J., Schut, A.F., Rutten, W.P., Pols, H.A., Witteman, J.C., Oostra, B.A., et al. (2007). Heritabilities, apolipoprotein E, and effects of inbreeding on plasma lipids in a genetically isolated population: the Erasmus Rucphen Family Study. Eur J Epidemiol 22, 99-105.

51. Weedon, M.N., and Frayling, T.M. (2008). Reaching new heights: insights into the genetics of human stature. Trends Genet 24, 595-603.

52. Hu, N., Wang, C., Hu, Y., Yang, H.H., Giffen, C., Tang, Z.Z., Han, X.Y., Goldstein, A.M., Emmert-Buck, M.R., Buetow, K.H., et al. (2005). Genome-wide association study in esophageal cancer using GeneChip mapping 10K array. Cancer Res 65, 2542-2546.

53. Thorleifsson, G., Walters, G.B., Gudbjartsson, D.F., Steinthorsdottir, V., Sulem, P., Helgadottir, A., Styrkarsdottir, U., Gretarsdottir, S., Thorlacius, S., Jonsdottir, I., et al. (2009). Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nat Genet 41, 18-24.

54. Dawber, T.R., Meadors, G.F., and Moore, F.E., Jr. (1951). Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health Nations Health 41, 279-281.

55. Hofman, A., Breteler, M.M., van Duijn, C.M., Krestin, G.P., Pols, H.A., Stricker, B.H., Tiemeier, H., Uitterlinden, A.G., Vingerling, J.R., and Witteman, J.C. (2007). The Rotterdam Study: objectives and design update. Eur J Epidemiol 22, 819-829.

56. Psaty, B.M., O'Donnell, C.J., Gudnason, V., Lunetta, K.L., Folsom, A.R., Rotter, J.I., Uitterlinden, A.G., Tamara B. Harris, Whitteman, J.C.M., and Boerwinkle, E. (2009). Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium Design of Prospective Meta-Analyses of Genome-Wide Association Studies From 5 Cohorts Circ Cardiovasc Genet 2, 73-80.

57. Stolk, R.P., Rosmalen, J.G., Postma, D.S., de Boer, R.A., Navis, G., Slaets, J.P., Ormel, J., and Wolffenbuttel, B.H. (2008). Universal risk factors for multifactorial diseases: LifeLines: a three-generation population-based study. Eur J Epidemiol 23, 67-74.

58. International Schizophrenia, C., Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., and Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460, 748-752.

59. Miettinen, O.S. (2009). Up from 'false positives' in genetic-and other-epidemiology. Eur J Epidemiol 24, 1-5.

60. Evans, D.M., Visscher, P.M., and Wray, N.R. (2009). Harnessing the Information Contained Within Genome-wide Association Studies to Improve Individual Prediction of Complex Disease Risk. Hum Mol Genet.

61. Janssens, A.C. (2008). Is the time right for translation research in genomics? Eur J Epidemiol 23, 707-710.

62. Drame, M., Novella, J.L., Lang, P.O., Somme, D., Jovenin, N., Laniece, I., Couturier, P., Heitz, D., Gauvain, J.B., Voisin, T., et al. (2008). Derivation and validation of a mortality-risk index from a cohort of frail elderly patients hospitalised in medical wards via emergencies: the SAFES study. Eur J Epidemiol 23, 783-791.

63. Janssens, A.C., Aulchenko, Y.S., Elefante, S., Borsboom, G.J., Steyerberg, E.W., and van Duijn, C.M. (2006). Predictive testing for complex diseases using multiple genes: fact or fiction? Genet Med 8, 395-400.

64. van Hoek, M., Dehghan, A., Witteman, J.C., van Duijn, C.M., Uitterlinden, A.G., Oostra, B.A., Hofman, A., Sijbrands, E.J., and Janssens, A.C. (2008). Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study. Diabetes 57, 3122-3128.

65. Demirkan, A., Penninx, B.W., Hek, K., Wray, N.R., Amin, N., Aulchenko, Y.S., van Dyck, R., de Geus, E.J., Hofman, A., Uitterlinden, A.G., et al. (2010). Genetic risk profiles for depression and anxiety in adult and elderly cohorts. Mol Psychiatry.

66. Ramdas, W.D., Amin, N., van Koolwijk, L.M., Janssens, A.C., Demirkan, A., de Jong, P.T., Aulchenko, Y.S., Wolfs, R.C., Hofman, A., Rivadeneira, F., et al. (2010). Genetic architecture of open angle glaucoma and related determinants. J Med Genet.

67. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42, 565-569.

68. Goldstein, D.B. (2009). Common genetic variation and human traits. N Engl J Med 360, 1696-1698.

69. Cirulli, E.T., and Goldstein, D.B. (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet 11, 415-425.

70. Kong, A., Steinthorsdottir, V., Masson, G., Thorleifsson, G., Sulem, P., Besenbacher, S., Jonasdottir, A., Sigurdsson, A., Kristinsson, K.T., Frigge, M.L., et al. (2009). Parental origin of sequence variants associated with complex diseases. Nature 462, 868-874.

71. Kong, A., Masson, G., Frigge, M.L., Gylfason, A., Zusmanovich, P., Thorleifsson, G., Olason, P.I., Ingason, A., Steinberg, S., Rafnar, T., et al. (2008). Detection of sharing by descent, long-range phasing and haplotype imputation. Nat Genet 40, 1068-1075.

72. Siva, N. (2008). 1000 Genomes project. Nat Biotechnol 26, 256.

73. Sobreira, N.L., Cirulli, E.T., Avramopoulos, D., Wohler, E., Oswald, G.L., Stevens, E.L., Ge, D., Shianna, K.V., Smith, J.P., Maia, J.M., et al. (2010). Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene. PLoS Genet 6, e1000991.

74. Walsh, T., Shahin, H., Elkan-Miller, T., Lee, M.K., Thornton, A.M., Roeb, W., Abu Rayyan, A., Loulus, S., Avraham, K.B., King, M.C., et al. (2010). Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82. Am J Hum Genet 87, 90-94.

[11]LifeLines Cohort Study & Biobank, University Medical Center Groningen, University of Groningen, the Netherlands
[12]Institute for Community Medicine, University of Greifswald, Germany
[13]Department of Prosthodontics, Gerodontology and Dental Materials, Center of Oral Health, University of Greifswald, Germany
[14]Institute of Clinical Chemistry and Laboratory Medicine, University of Greifswald, Germany
[15]Department of Psychiatry, University of Mainz, Mainz, Germany
[16]National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA
[17]Center for Population Studies, NHLBI, Bethesda, Maryland, USA
[18]Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA, USA
[19]Group Health Research Institute, Group Health Cooperative, Seattle, WA, USA
[20]Icelandic Heart Association, Kopavogur, Iceland
[21]University of Iceland, Reykjavik, Iceland
[22]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland, USA
[23]Department of Epidemiology and Medicine, Johns Hopkins University, Baltimore, Maryland, USA
[24]deCODE Genetics, Reykjavik, Iceland
[25]Department of internal medicine, Erasmus University Medical Center, Rotterdam, the Netherlands
[26]Human Genetics, Wellcome Trust Sanger Institute, Genome Campus, Hinxton, UK
[27]Department of Child and Adolescent Psychiatry, Erasmus University Medical Center, Rotterdam, The Netherlands
[28]Department of Endocrinology, University Medical Center Groningen, University of Groningen, the Netherlands
[29]Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, the Netherlands
[30]Department of Psychiatry, Washington University, St Louis, USA
[31]Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany
[32]Department of Psychiatry and Psychotherapy, University of Greifswald, Stralsund, Germany
[33]Centre of Medical Systems Biology, Netherlands Consortium on Health Aging and National Genomics Initiative

* Authors contributed equally

Chapter 5

General Discussion

Findings of this thesis

Behaviour refers to the actions of an organism, usually in relation to its environment. In humans, behaviour appears to be controlled primarily by the endocrine and nervous systems. The complexity of behaviour in humans is correlated to the complexity of the nervous system. Behaviours can be innate or learned. Many behaviours are highly heritable both within and across species.

In this thesis I have made an effort to unfold the genetics of traits that have a strong behavioural aspect and are known to affect social, somatic and psychological health. The traits I have studied concern a variety of outcomes including personality, ADHD, coffee drinking and sleep. All are common and/or quantitative traits and also known to be significantly heritable. I have made use of various gene mapping techniques including linkage and association. Moreover, I also use more advanced molecular and statistical analysis including copy number variation (CNV) and genomic imprinting analysis to find genetic variants associated with these traits. Using these methods I discovered novel loci (genes) implicated in the traits studied.

The first trait studied in this thesis is attention deficit hyperactivity disorder (ADHD). ADHD is a psychiatric developmental disorder[1] with an extremely strong behavioural component, which results in impaired social and/or academic functioning. In **chapter 2** we report the findings of a genome-wide linkage analysis of ADHD in nine inbred children identified in the Genetic Research in Isolated Populations (GRIP program). We found suggestive evidence of linkage (LOD = 2.2) of ADHD to a novel locus at 18q21-22 using homozygosity mapping. This finding was also supported by the genome-wide association analysis performed by the Genetic Association Information Network (GAIN) (rs2311120, p-value = $1.06*10^{-05}$). In a sub-analysis of inattentive type ADHD children we also found suggestive evidence of linkage to 6p22 (LOD=2), a region which was found earlier for ADHD in patients with reading disability[2]. Other regions with HLOD ≥ 1 include 1p36, 5q33, 6p12, 18p11, and 15q25. Although not significant, most of these regions including 1p36, 5q33 and 6p12 overlap or are adjacent to those found by the previous studies[3-5] and harbour serotonin receptors including *HTR1B, HTR1E, HTR4, HTR1D,* and *HTR6*. This suggests that linkage analysis is a powerful approach for identification of genetic loci as we were able to discover several interesting candidate regions with only nine patients compared to the 958 trios in the GAIN analysis.

I next focused on personality traits. The structure of human personality has traditionally been accounted for by a relatively small set of traits. Over the last century, scientific consensus has converged on a taxonomic model of personality traits based on five higher-order dimensions of Neuroticism, Extraversion, Openness to Experience, Agreeableness and Conscientiousness, known as the Five-Factor Model (FFM)[6]. These five dimensions are developed as orthogonal scales and therefore are largely independent. Together they provide a broad description of personality. Neuroticism is commonly defined as emotional instability; it involves the experience of negative emotions such as anxiety, depression, hostility, and

vulnerability to stress. Extraversion is characterized by positive emotions, gregariousness, and the tendency to be active, seek out stimulation and enjoy the company of others. Openness to Experience involves active imagination, aesthetic attentiveness, variety preference and intellectual curiosity. Agreeableness can be defined as the tendency to be cooperative and compassionate rather than suspicious and antagonistic towards others. Lastly, the dimension of Conscientiousness reflects self-discipline, carefulness, thoroughness, organization, deliberation and achievement. Personality traits predict a host of social, behavioural and health outcomes, such as job performance, longevity, and many psychiatric disorders, including substance abuse and dependency, mood disorders such as major depressive disorder (MDD), anxiety disorders, and personality disorders[7-20]. Although highly heritable with heritability estimates ranging between 33% and 65%[21-25], finding genes implicated in personality has proved to be difficult.

During my PhD training, I participated in a large genome-wide association study (GWA) aiming to identify common genetic variants with small effects. Included in the analyses were ten discovery samples (17,375 adults) and five in-silico replication samples (3,294 adults). All participants were of European ancestry. Personality scores for Neuroticism, Extraversion, and Openness to experience, Agreeableness, and Conscientiousness were based on the NEO Five-Factor Inventory (NEO-FFI)[6]. Genotype data were available of ~ 2.4 million single nucleotide polymorphisms (SNPs; both directly typed and imputed using HAPMAP data). In the discovery samples, classical association analyses were performed under an additive model followed by meta-analysis using the weighted inverse variance method. Results showed genome-wide significance for Openness to experience near the *RASA1* gene on 5q14.3 (rs1477268 and rs2032794, p-value = $2.8*10^{-8}$ and p-value = $3.1*10^{-8}$) and for Conscientiousness in the brain-expressed *KATNAL2* gene on 18q21.1 (rs2576037, p-value = $4.9*10^{-8}$). We further conducted a gene-based test that confirmed the association of *KATNAL2* to Conscientiousness. In-silico replication did not, however, show significant associations of the top SNPs with Openness and Conscientiousness, although the direction of effect of the *KATNAL2* SNP on Conscientiousness was consistent in all replication samples. Larger scale GWA studies and alternative approaches are required for confirmation of *KATNAL2* as a novel gene affecting Conscientiousness.

For this thesis I followed an alternative approach. To find genetic loci that may confer moderate to large effects on personality traits of Neuroticism, Extraversion, Openness, Agreeableness and Conscientiousness, in **chapter 3.1** I performed qualitative genome-wide linkage scans of high scoring individuals (> 90^{th} percentile) for each of the five traits in the Erasmus Rucphen Family (ERF) Study. We found significant evidence of linkage of Conscientiousness to 20p13 under the recessive model (log of odds (LOD) = 5.86). Haplotype construction revealed a distinct haplotype segregating in six high LOD score contributing families, which showed marginal association (p-value = 0.058 and 0.038) under a recessive model. This haplotype covers a 100 kilobase (kb) region on 20p13 and contains 5 genes belonging to the beta defensin family (*DEFB*). We also found suggestive evidence of linkage of Neuroticism to 21q22 (LOD = 3.42),

where we found two significant haplotypes (p-values = 0.009, 0.007), one of which was being shared in a homozygous state by most of the high scoring individuals from three high LOD score contributing families. This region harbors 10 genes, including *TTC3, DSCR9, DSCR3, DYRK1A, KCNJ6* and *PIGP*, and has been associated with cognitive and mental impairment in individuals with Down's syndrome[26]. Suggestive linkage (LOD = 3.32) yet a significant haplotype (p-value = 0.018) was found for Agreeableness in the 17q24 region. Genes in this region include *KCNJ2, KCNJ16* and *MAP2K6*. Excess homozygosity in each of these linked regions prompted us to look for deletions in the regions. However, the copy number variation (CNV) analysis did not provide evidence for CNVs. These regions did not emerge in the GWAS I mentioned earlier. There may be several explanations for this including a false positive finding in this study. However, more likely, GWA studies are unlikely to pick up relatively rare recessive variants[27]. Such variants may indeed explain part of the missing heritability in quantitative traits[27].

In **chapter 3.2** we performed an alternative linkage analysis including the various personality outcomes quantitatively. Although quantitative traits are a powerful approach, the number of subjects to be studied is critical. Further, also for linkage analysis replication is important to exclude false positive findings. Therefore, a meta-analysis of four un-published quantitative trait linkage analyses of the five personality traits was conducted. The samples (*N* = 6,149) included Erasmus Rucphen Family Study (ERF), Netherlands Twin Register (NTR) and an adult and adolescent sample from Australia (QIMR). To maximize power we combined genome-wide linkage and association analyses methods in a two step analysis [28]. The regions of significant and suggestive evidence of linkage were fine mapped using the results of the largest meta-analysis of genome-wide association studies (*N* ~ 18,000) of these five personality traits. Significant linkage was detected for Neuroticism at chromosome 3p14 (LOD = 4.67) and at chromosome 19q13 (LOD = 3.55), for Extraversion at 14q32 (LOD = 3.3); for Agreeableness at 3p25 (LOD = 3.67) and at two adjacent regions on chromosome 15 including 15q13 (LOD = 4.07) and 15q14 (LOD = 3.52) in the individual scans. In the meta-analysis, we found strong evidence of linkage of Extraversion to 4q34, 9q34, 10q24 and 11q22, Openness to 2p25, 3q26, 9p21, 11q24, 15q26 and 19q13, and Agreeableness to 4q34 and 19p13. When combining these data with the association results of the GWAS of these personality traits, significant evidence of association was detected for Openness at 11q24 (rs677035, p-value = $2.6*10^{-06}$). Borderline evidence for association was detected between Neuroticism and rs332389 (p-value = $9*10^{-05}$) at 3p14 and between Extraversion and rs7088779 (p-value = $4.2*10^{-06}$) at 10q24. Of these three SNPs, rs677035 is an intergenic SNP located between *FLI1* and *KCNJ1*, rs332389 is intronic to the gene *SLC25A26 a*nd rs7088779 is located between *CRTAC1* (cartilage acidic protein 1) and *C10orf28*; a region previously implicated in Alzheimer's disease. Also in the quantitative linkage analysis, which is more similar to GWAS, there was no evidence for a role of *KATNAL2* or *RASA1* for Conscientiousness and Openness. Again one may argue that this may be explained by a false positive finding or the differences in power for the various variants implicated in these

complex traits. An interesting gene family that emerges from the linkage studies in this thesis is the KCNJ family. The linkage regions we identified using both dichotomous and quantitative outcome are enriched with KCNJ genes including *KCNJ6, KCNJ16, KCNJ2* and *KCNJ1*. KCNJs are inwardly rectifying potassium channels. They are the targets of multiple toxins, and malfunction of the channels has been implicated in several diseases including neurological and endocrine disorders[29] and electrocardiogram arrhythmias[30].

Epigenetic effects, of which genomic imprinting is a form, is often cited as a reason for the loss of efficiency of genome-wide association studies in identifying common variants for complex traits[31]. In **chapter 3.3** we performed a genome-wide search for imprinted regions for personality traits in the ERF sample using SNPs from the Illumina 6K linkage panel. We found significant differences in paternally and maternally transmitted alleles at 6q25 (rs602890, p-value = $5*10^{-07}$) for Extraversion and at 4q21 (rs1880719, p-value = $8*10^{-07}$) for Agreeableness. The two SNPs, which did not show any significance in a general association test, showed significant association when parental origin of alleles was incorporated in the association model (p-value= $3*10^{-06}$ for rs602890 and p-value = $1*10^{-06}$ for rs1880710). Rs602890 showed opposite and nominally significant effects when inherited from father (β = -2.78, p-value = $2.7*10^{-03}$) and mother (β = 4.33; p-value = $2*10^{-04}$) and rs1880719 showed significant association with Agreeableness only when inherited from the mother (β = -4.47, p-value = $6*10^{-05}$). The 6q25 region is a known imprinted region[32], while imprinting is also hypothesized as the reason for the central nervous system overgrowth in the 4q21/4q23 syndrome[33]. Rs602890 is an intronic SNP in the gene *ZDHHC14* and rs1880719 is an intronic SNP in the gene *SCD5* encoding Stearoyl-CoA desaturase. None of these SNPs showed any evidence of association in the large GWAS or linkage studies when ignoring parent-of-origin effects. These findings can only be replicated in a family based sample. An attempt to replicate the two most interesting findings (rs1880719 and rs602890) in the Australian twin samples failed to provide sufficient evidence of parent-of-origin effects. We assume that this may be due to the sparse set of markers that we used in our analysis (5,250), making it difficult to pinpoint the region of interest precisely. A fine mapping of the 4q21 region revealed four additional significant SNPs (in high LD with rs1880719) in the same gene (*SCD5*). At present another replication effort is being made in the twin sample from the Netherlands Twin Registry (NTR).

Finally, we have studied sleep and caffeine in the thesis. Epidemiological studies have shown that sleep duration is associated with psycho-somatic malfunctioning[34-37]. In **chapter 4.2** we performed the first large scale genome-wide association analysis of sleep duration in 7 European cohorts (N = 4,260). We identified an intronic variant on chromosome 12p12.1 (rs11046205) in the *ABCC9* gene and validated it in a replication study. When the conserved *Drosophila* homologue was knocked down pan-neuronally by RNAi, night sleep duration was reduced. *ABCC9* encodes an ATP-sensitive potassium channel subunit (*SUR2*), which functions as a sensor of intracellular energy. Our findings therefore link sleep and its restorative effects,

known in physiological terms as the homeostat, to cellular energy metabolism. Time to fall asleep (sleep latency) is a major determinant of the quality of sleep. Increased sleep latency is one of the major characteristic of sleep onset insomnia[38] and delayed sleep phase syndrome[39], which are two common circadian disorders that hamper an individual's social and academic life and work performance. With no previous genetic studies performed on this trait, in **chapter 4.1** we performed a meta-analysis of seven genome-wide association studies consisting of 4,270 individuals of Caucasian origin. We found strong evidence of association of a variant (rs9900428, p-value = $5.7*10^{-08}$) of the hexaribonucleotide binding protein 3 (*HRNBP3*) (also known as *FOX3*) with sleep latency. We successfully replicated our findings in five additional cohorts ($N = 10,901$). The meta-analysis of discovery and replication cohorts showed a genome-wide significant association of rs9900428 (p-value = $3.8*10^{-08}$). The *FOX3* gene is a highly brain specific protein that belongs to the family of FOX genes, which are regulators of pre and postnatal alternative splicing in muscle and brain tissues. *FOX3* is suggested to be an upstream regulator for alternative splicing of brain specific mRNAs involved in glutamatergic and GABAergic transmission. This is the first genome search on sleep latency and no variant for sleep latency has been reported earlier.

We finally study a major determinant of sleep latency, caffeine. Caffeine is known to have behavioural effects when taken in moderate amounts[40-42]. High doses of caffeine can have also negative effects such as anxiety, restlessness and insomnia. Caffeine intake is also described as a model of drug abuse[43] and coffee is the major source of caffeine in adults. In **chapter 4.3** we performed a large scale meta-analysis of genome-wide association studies on coffee intake. Our data ($N \sim 18,000$) consisted of samples from eight cohorts. We found genome-wide significant evidence of association of a variant (rs2470893) in 15q25 region, intergenic to genes *CYP1A1* and *CYP1A2*. The "T" allele of rs2470893 was positively associated with coffee drinking. The *CYP1A1* gene was also found to be down-regulated after caffeine treatment in lymphoblastoid cell lines. *CYP1A1* is known to metabolize polycyclic aromatic hydrocarbons such as benzo(a)pyrene, which is an important constituent of coffee and also known to be involved in the mutagenesis resulting from coffee intake while *CYP1A2* is involved in primary metabolism of caffeine. Combined evidence from GWAS and caffeine expression analysis suggests that *CYP1A1* regulates coffee drinking and that this association is independent of smoking status. Apart from *CYP1A1/CYP1A2* we also found strong evidence of association of coffee intake with polymorphisms in *NCALD* (rs16868941), *NRCAM* (rs382140) and *CAB39L* (rs9526558) genes. Earlier we found strong association of the *NCALD* gene with sleep latency (rs17498920, p-value = $2*10^{-05}$). This is an interesting finding since coffee intake is known to interfere with melatonin secretion and delay the onset of sleep. A replication analysis of the top findings from the GWAS in a further 8,000 individuals of Dutch origin (LifeLines), confirmed association of *CYP1A1/A2* locus and rs382140 near the *NRCAM* gene with coffee drinking. *CAB39L*, although not successfully replicated was found to be up-regulated in all the three cell lines studied after caffeine treatment suggesting

that the gene is indeed involved in caffeine metabolism and intake. Association analysis of the polymorphisms in these genes with coffee related phenotypes including blood pressure, mortality, cancer mortality, lipids and Alzheimer disease (AD) revealed significant association of rs2470893 (*CYP1A1/1A2*) with systolic and diastolic blood pressure, where again "T" allele was positively associated, suggesting that increased coffee intake elevates blood pressure[44]. The association of our top hits with blood pressure is very interesting and strengthens the argument that coffee consumption elevates blood pressure[45].

In search of the missing heritability

For the past 5 years genome-wide association studies (GWAS) have dominated the search for new genes for complex diseases overtaking other approaches of gene finding such as candidate gene and linkage analyses. Facilitated by technological developments in molecular biology, genetic epidemiologists have so far discovered many variants associated with several common diseases and traits such as Type 2 Diabetes, age-related macular degeneration and Crohn's disease[46]. There currently are 26 established susceptibility genes published for type 2 diabetes[47] 54 for human height and 22 for lipid levels[48-49]. These variants still explain only a small part of the genetic variance or heritability, for human height and lipids up to 4–6%[50-51], and subsequently the search for novel variants continues to unravel 'missing heritability'.

This missing heritability is explained by additional rare variants with strong effects and/or common variants with weak effects, acting additively and/or interacting with other genetic and environmental variants. To discover these additional genetic factors, GWAS need to enlarge, and this has led to further expansion of existing consortia and the establishment of new ones. Since the first publication in 2005[52], GWAS have undergone enormous evolution: from 10,000 single nucleotide polymorphisms (SNPs) in 100 individuals of a single sample[52] to 1 million genotyped and ~ 2.5 million imputed SNPs in more than 80,000 individuals of multiple samples [53]. The decreasing costs of genotyping, new statistical methodologies, and increasing willingness of the scientists to share and pool data sets have facilitated these rapid developments and made this approach very successful also in the setting of epidemiology. For instance, the Cohort for Health and Aging Research (CHARGE) is studying multiple common traits in 50,000–70,000 individuals from US and European follow-up studies[54-56], and the Dutch three-generation study LifeLines is going to include 165,000 participants[57]

While increasing size will help in finding new variants with smaller effects, there will also be true positives that remain undetected in the larger consortia because of the stringent threshold levels of statistical significance imposed in GWAS (p-value < 5 *10^{-08}) to adjust for multiple testing. The chances of success of consortia are further reduced if confounding due to population heterogeneity, also refer to as population admixture, is to be adjusted for, which is the case when populations are of different genetic origins. Therefore new approaches are needed to identify genetic variants explaining the missing heritability and one such new approach was

used successfully in a recent GWAS in schizophrenia that was published in *Nature*[58]. The classical GWAS analysis produced only one genome-wide significant polymorphism, but the authors used a new 'genetic scoring' method through which they demonstrated that there indeed existed undetected variants below the threshold. How to detect variants that are not detected? Basically, the method tests the association of a score variable that manifests a combined effect of many SNPs. The polymorphisms in the score are selected on the basis of their nominal p-value in the predefined discovery sample. Scores can be generated for any arbitrarily chosen threshold of nominal statistical significance, for instance selecting all SNPs with p-values lower than e.g. 0.01, 0.1 or 0.5. The significance of the score is then tested by using it as a predictor in a simple regression model in an independent 'target sample'. In this target sample, a one-parameter test for all SNPs can be used, thus relaxing the conservative p-value of $5*10^{-08}$ needed for testing all SNPs in GWAS to classical significance level of 0.05. Using data from the International Schizophrenia Consortium with men defined as the discovery sample and women as the target sample, the authors showed that a score based on all SNPs with p-value < 0.5 was most strongly and significantly correlated with schizophrenia in the target sample compared to the scores based on other thresholds. The fact that the set of SNPs with p-value < 0.5, including both many falsely and an unknown number of truly associated SNPs, predicted better than the score with p-value < $5*10^{-08}$ suggests that both the number of undetected relevant variants as well as their joint effect on the outcome is substantial[58]. The authors further showed that the score correlated significantly with related diseases as bipolar disorder, but not with unrelated outcomes such as Crohn's disease, coronary artery disease, hypertension, rheumatoid arthritis or type 1 and type 2 diabetes. This suggests that schizophrenia and bipolar disorder have a shared genetic component and also that the selected alleles were specific to schizophrenia and related disorders[58].

The genetic scoring method is logical and simple as among the SNPs that fail to reach the significance threshold in the GWAS there ought to be true associations, which just do not reach the threshold because the study does not have enough power[59]. There may, however, be several caveats. First, the informative value of the approach depends on the size of the discovery sample. If the discovery sample is small, more falsely associated SNPs will be selected at each threshold, and consequently scores do not explain much of the phenotypic variance in the target sample. The second caveat is that also a score based on 38,000 SNPs with a p-value lower than 0.5, derived in a discovery sample of 3,800 individuals, explained only 3% of the heritable variance in the target population of 3,100 persons. It can be expected that a larger discovery set will select more true positives among those with a p-value lower than 0.5 and therefore explain a higher percentage of the variance in the target sample. However, simulations showed that the variance explained by the scores can increase from 3 to 20% if the size of the discovery sample is increased to 20,000 individuals[58]. Thus, also for this new method the size of the discovery sample is an important determinant of success. Third, one of the major conclusions on the basis of this

method is that there are undetected common genetic contributions. Of course one may argue that this observation could already be inferred from the fact that there is 'missing heritability'. But perhaps an even more important limitation of the genetic scoring method is that it does not tell which one(s) of the variants included is responsible for the statistical significance.

Then what can we do with this information? First, the method may be used to improve our understanding the genetic architecture of the disease or trait. Scores can be calculated and tested for multiple different significance thresholds levels of statistical significance. By comparing the proportions of explained variance across these thresholds, a pattern may be observed. When going up from a very low threshold, e.g., p-value < 10^{-07} to p-value < 0.5, we may see that scores may rise to a certain point and then either decline or become stable, a pattern which suggests that a few genes with stronger effects may be involved. When the proportion of explained variance monotonically increases until all SNPs are included in the scores, there are likely to be a large number of common variants with small effects. So the scores calculated over several different cut offs can give an indication on how complex the trait is, on the likelihood that the trait has a polygenic basis. For example, for schizophrenia the score goes up from 0.004 to 0.025 by moving up from a threshold of p-value < 0.01 to p-value < 0.5[58], which is an indication that many more common low risk variants are likely involved in schizophrenia.

Second, this method could be considered as an intermediate step in the gene discovery process. When scores are statistically significant, one may consider to only analyzing the included SNPs in the independent samples. For replication purposes this leads to a less stringent level of statistical significance, and potentially to a higher likelihood of finding susceptibility variants. Because the success of this approach will depend on the size of the discovery sample — the larger the discovery sample the more likely true susceptibility genes will be selected in the scores — its added value of selecting SNPs in much smaller independent populations may not be efficient. More promising is to use the score approach to select SNPs for use in complex modelling of the trait for instance to study gene by gene interactions which otherwise seems impossible with 2.5 million SNPs.

Third, the method could be used to predict disease for preventive and clinical purposes. Evans and colleagues applied the score approach and assessed the discriminative ability for several threshold levels of statistical significance in several complex diseases [60]. When significance thresholds were varied from 10^{-05} to 0.8, discriminative ability improved for bipolar depression, coronary heart disease, hypertension and type 2 diabetes, but decreased for rheumatoid arthritis and type 1 diabetes prediction. For all diseases, the discriminative ability was lower than what would be obtained when testing known susceptibility genes, except for hypertension where no susceptibility variants were known at the time and for bipolar disorder for which the score performed better than the known variants, but only for the liberal and not for the stringent significance thresholds. It is also suggested that the shared genetic liability between schizophrenia and bipolar disorder would make the genetic based refinement of the

diagnosis of these diseases possible, which may also be tried for other diseases with overlapping symptoms[58]. Of all potential applications of the genetic scoring method, this is the least substantiated, and it may still be too premature[61]. However, when the proportion of variance explained by the scores can increase from 3 to 20%, as suggested by simulation analyses by improving the power of the discovery set[58], the discriminative accuracy could be in the range of what we commonly see for non-genetic risk prediction models in e.g., cardiovascular diseases, diabetes and mortality[62-64].

GWAS have been very successful in finding multiple variants for many traits, but we are reaching the limits of what can be found through this approach sooner or later. Whether the genetic scoring method will be successful in finding more risk variants for complex traits and in unravelling 'missing heritability' remains to be determined. So far we applied this method successfully in depression[65], lipids and glaucoma[66]. Although the approach works, the variance explained is still very low, limiting the value for application in public health or decision making. More recently a new approach has been proposed which uses the restricted maximum likelihood modelling (REML) to estimate the variance explained by the SNPs[67]. The authors show that using REML and kinships estimated from the genotyped SNPs, they could explain roughly half of the heritability of human height using common SNPs. For the remaining half they suggest incomplete LD between causal variant and the SNPs as the major limitation, which, they suggest, might be overcome using larger discovery samples. An issue to be evaluated is the extent to which the data over fits the regression model, limiting the predictive value to the population studied. It is clear that such over fitting makes the approach useless in the field of epidemiology or public health. The new genetic score method is one approach, approaches aiming at testing of more complex models with gene by gene and gene by environment interactions may be another avenue, although it is clear that epistasis and gene-environment interactions do not form a part of the missing heritability. Last but not least technological developments may come to rescue with new development in high throughput sequencing.

Future research

With whole-genome sequencing of hundreds of individuals within our reach, contemporary thinking in genetic epidemiology is distributed within two extremes: continue enlarging the GWAS to find more variants using the SNPs or to go forward with whole genome/exome sequencing with a view to find rare variants. The proponents of the former suggest that given enough sample size the GWAS will be able to find most of the genetic variants[67] associated with a certain trait or in other words people who strictly adhere to the common disease/common variant theory. Others suggest that we have reached the limit with the GWAS[68-69] and profess sequencing. As expected and based upon population genetics, my thesis shows that the two lines of thought are not mutually exclusive and should converge with infinitively large samples. We have witnessed overlap of our linkage findings for ADHD and personality traits with those of the respective GWAS.

If larger GWAS are to be performed, more complex modelling including parent-of-origin effects may be useful as we and others have shown[70]. Currently only family based studies are capable of performing such analysis. A limitation of these studies is that the replication samples are usually not available. However, more recently developed long range phasing method[71] may enable population based studies to incorporate parent-of-origin effects in the association models. Also for the complex traits, reducing the complexity of the trait, that is, the use of simpler base level phenotypes is another approach which can improve the performance/power of the GWAS to detect common variants. This is already being done for a variety of complex traits/disorders. For instance, each of the five personality traits mentioned earlier in this thesis are defined by six underlying sub-phenotypes. We are now performing the GWAS on each of the individual sub-phenotypes in an ongoing study. Expansion of the GWAS in terms of the number of SNPs used is yet another avenue which might lead to the discovery of more of the common variation. Most consortia are now switching to 1000 genomes[72] as the reference set for SNP imputations rather than the HapMap. This will increase the number of SNPs from 2.5 to 8 million to be used in the GWAS, thus improving the chances of finding more common variants.

Needless to say that whole genome/exome sequencing will reveal the rare variants in the genome, however, the enthusiasm is also coupled with a feeling of anxiety about what sequencing will reveal and whether or not we are capable of handling such an enormous amount of information with contemporary bioinformatics and methodological tools. Provided the technological and methodological advances, the ultimate goal is to do a whole genome sequencing of all the enrolled subjects and to perform genome-wide association studies of the 3 billion base pairs[69], meanwhile designs that are optimized to detect rare variants in smaller samples are suggested[69]. These include sequencing affected individuals in families and extreme-trait sequencing. Targeted sequencing of interesting regions might be something to start with. So far linkage studies have discovered several regions for many psychiatric and non-psychiatric traits. These regions are potentially interesting targets for sequencing and might reveal rare

mutations[73]. Homozygosity mapping may also be used as a tool to provide interesting regions for targeted sequencing[74].

The findings in this thesis especially the linkage findings are interesting targets for deep sequencing. Apart from these there are a number of findings in this thesis that require a follow up. First, is the KCNJ gene family that is linked to personality traits. These genes should be characterized in more detail in humans e.g. by sequencing. These genes may also link personality to related traits, for instance, heart and electrocardiogram (ECG) traits. For Sleep, the genes that we have identified (*ABCC9* for sleep duration and *HRNBP3*/*FOX3*) also require further investigation. Although, we successfully authenticate our findings by either performing gene knockout study or through replicating in other populations, the biological mechanisms through which these genes influence sleep traits remain elusive and need to be explored. For the genes that we have discovered for coffee intake, it will be of interest to link the findings to coffee related traits and diseases. Another interesting avenue will be to look at these genes from the perspective of addiction.

REFERENCES

1. Zwi, M., Ramchandani, P., and Joughin, C. (2000). Evidence and belief in ADHD. BMJ 321, 975-976.
2. Willcutt, E.G., Pennington, B.F., Smith, S.D., Cardon, L.R., Gayan, J., Knopik, V.S., Olson, R.K., and DeFries, J.C. (2002). Quantitative trait locus for reading disability on chromosome 6p is pleiotropic for attention-deficit/hyperactivity disorder. Am J Med Genet 114, 260-268.
3. Arcos-Burgos, M., Castellanos, F.X., Pineda, D., Lopera, F., Palacio, J.D., Palacio, L.G., Rapoport, J.L., Berg, K., Bailey-Wilson, J.E., and Muenke, M. (2004). Attention-deficit/hyperactivity disorder in a population isolate: linkage to loci at 4q13.2, 5q33.3, 11q22, and 17p11. Am J Hum Genet 75, 998-1014.
4. Zhou, K., Asherson, P., Sham, P., Franke, B., Anney, R.J., Buitelaar, J., Ebstein, R., Gill, M., Brookes, K., Buschgens, C., et al. (2008). Linkage to chromosome 1p36 for attention-deficit/hyperactivity disorder traits in school and home settings. Biol Psychiatry 64, 571-576.
5. Ogdie, M.N., Fisher, S.E., Yang, M., Ishii, J., Francks, C., Loo, S.K., Cantor, R.M., McCracken, J.T., McGough, J.J., Smalley, S.L., et al. (2004). Attention deficit hyperactivity disorder: fine mapping supports linkage to 5p13, 6q12, 16p13, and 17p11. Am J Hum Genet 75, 661-668.
6. Costa, P.T., and McCrae, R.R. (1992). Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI) Professional Manual. Psychological Assessment Resources.
7. Bienvenu, O.J., Hettema, J.M., Neale, M.C., Prescott, C.A., and Kendler, K.S. (2007). Low extraversion and high neuroticism as indices of genetic and environmental risk for social phobia, agoraphobia, and animal phobia. Am J Psychiatry 164, 1714-1721.
8. Budaev, S.V. (1999). Personality and Individual Differences. Elsevier 26, 801-813.
9. Cramer, D. (1991). Type A behaviour pattern, extraversion, neuroticism and psychological distress. Br J Med Psychol 64 (Pt 1), 73-83.
10. Crowe, M., Andel, R., Pedersen, N.L., Fratiglioni, L., and Gatz, M. (2006). Personality and risk of cognitive impairment 25 years later. Psychol Aging 21, 573-580.
11. Distel, M.A., Trull, T.J., Willemsen, G., Vink, J.M., Derom, C.A., Lynskey, M., Martin, N.G., and Boomsma, D.I. (2009). The five-factor model of personality and borderline personality disorder: a genetic analysis of comorbidity. Biol Psychiatry 66, 1131-1138.
12. Goodwin, R.D., Castro, M., and Kovacs, M. (2006). Major depression and allergy: does neuroticism explain the relationship? Psychosom Med 68, 94-98.
13. Kendler, K.S., Gatz, M., Gardner, C.O., and Pedersen, N.L. (2006). Personality and major depression: a Swedish longitudinal, population-based twin study. Arch Gen Psychiatry 63, 1113-1120.
14. Meins, W., and Dammast, J. (2000). Do personality traits predict the occurrence of Alzheimer's disease? Int J Geriatr Psychiatry 15, 120-124.
15. Menza, M. (2000). The personality associated with Parkinson's disease. Curr Psychiatry Rep 2, 421-426.
16. Robertson, D.A., Ray, J., Diamond, I., and Edwards, J.G. (1989). Personality profile and affective state of patients with inflammatory bowel disease. Gut 30, 623-626.
17. Sutin, A.R., Scuteri, A., Lakatta, E.G., Tarasov, K.V., Ferrucci, L., Costa, P.T., Jr., Schlessinger, D., Uda, M., and Terracciano, A. (2010). Trait antagonism and the progression of arterial thickening: women with antagonistic traits have similar carotid arterial thickness as men. Hypertension 56, 617-622.
18. Sutin, A.R., Terracciano, A., Deiana, B., Naitza, S., Ferrucci, L., Uda, M., Schlessinger, D., and Costa, P.T., Jr. (2010). High neuroticism and low conscientiousness are associated with interleukin-6. Psychol Med 40, 1485-1493.
19. Sutin, A.R., Terracciano, A., Deiana, B., Uda, M., Schlessinger, D., Lakatta, E.G., and Costa, P.T., Jr. (2010). Cholesterol, triglycerides, and the Five-Factor Model of personality. Biol Psychol 84, 186-191.
20. Terracciano, A., Lockenhoff, C.E., Crum, R.M., Bienvenu, O.J., and Costa, P.T., Jr. (2008). Five-Factor Model personality profiles of drug users. BMC Psychiatry 8, 22.
21. Jang, K.L., Livesley, W.J., and Vernon, P.A. (1996). Heritability of the big five personality dimensions and their facets: a twin study. J Pers 64, 577-591.

22. Jang, K.L., McCrae, R.R., Angleitner, A., Riemann, R., and Livesley, W.J. (1998). Heritability of facet-level traits in a cross-cultural twin sample: support for a hierarchical model of personality. J Pers Soc Psychol 74, 1556-1565.

23. Loehlin, J.C., McCrae, R.R., Paul T. Costa, j., and John, O.P. (1998). Heritabilities of common and measure-specific components of the big five personality factors. Journal of research in personality 32, 431-453.

24. Riemann, R., Angleitner, A., and Strelau, J. (1997). Genetic and environmental influences on personality: A study of twins reared together using the self- and peer report NEO-FFI scales. J Personal 65, 449-475.

25. Van Os, J., and Jones, P.B. (2001). Neuroticism as a risk factor for schizophrenia. Psychol Med 31, 1129-1134.

26. Altafaj, X., Dierssen, M., Baamonde, C., Marti, E., Visa, J., Guimera, J., Oset, M., Gonzalez, J.R., Florez, J., Fillat, C., et al. (2001). Neurodevelopmental delay, motor abnormalities and cognitive deficits in transgenic mice overexpressing Dyrk1A (minibrain), a murine model of Down's syndrome. Hum Mol Genet 10, 1915-1923.

27. Maher, B. (2008). Personal genomes: The case of the missing heritability. Nature 456, 18-21.

28. Sham, P.C., Cherny, S.S., Purcell, S., and Hewitt, J.K. (2000). Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. Am J Hum Genet 66, 1616-1630.

29. Abraham, M.R., Jahangir, A., Alekseev, A.E., and Terzic, A. (1999). Channelopathies of inwardly rectifying potassium channels. FASEB J 13, 1901-1910.

30. Priori, S.G., Pandit, S.V., Rivolta, I., Berenfeld, O., Ronchetti, E., Dhamoon, A., Napolitano, C., Anumonwo, J., di Barletta, M.R., Gudapakkam, S., et al. (2005). A novel form of short QT syndrome (SQT3) is caused by a mutation in the KCNJ2 gene. Circ Res 96, 800-807.

31. Eichler, E.E., Flint, J., Gibson, G., Kong, A., Leal, S.M., Moore, J.H., and Nadeau, J.H. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet 11, 446-450.

32. Paterson, A.D., Naimark, D.M., and Petronis, A. (1999). The analysis of parental origin of alleles may detect susceptibility loci for complex disorders. Hum Hered 49, 197-204.

33. Nowaczyk, M.J., Teshima, I.E., Siegel-Bartelt, J., and Clarke, J.T. (1997). Deletion 4q21/4q22 syndrome: two patients with de novo 4q21.3q23 and 4q13.2q23 deletions. Am J Med Genet 69, 400-405.

34. Taheri, S., and Thomas, G.N. (2008). Is sleep duration associated with obesity-where do U stand? Sleep Med Rev 12, 299-302.

35. Stickgold, R. (2001). Toward a cognitive neuroscience of sleep. Sleep Med Rev 5, 417-421.

36. Gangwisch, J.E. (2009). Epidemiological evidence for the links between sleep, circadian rhythms and metabolism. Obes Rev 10 Suppl 2, 37-45.

37. Ikehara, S., Iso, H., Date, C., Kikuchi, S., Watanabe, Y., Wada, Y., Inaba, Y., Tamakoshi, A., and Group, J.S. (2009). Association of sleep duration with mortality from cardiovascular disease and other causes for Japanese men and women: the JACC study. Sleep 32, 295-301.

38. Eddy, M., and Walbroehl, G.S. (1999). Insomnia. Am Fam Physician 59, 1911-1916, 1918.

39. Weitzman, E.D., Czeisler, C.A., Coleman, R.M., Spielman, A.J., Zimmerman, J.C., Dement, W., Richardson, G., and Pollak, C.P. (1981). Delayed sleep phase syndrome. A chronobiological disorder with sleep-onset insomnia. Arch Gen Psychiatry 38, 737-746.

40. Mitchell, P.J., and Redman, J.R. (1992). Effects of caffeine, time of day and user history on study-related performance. Psychopharmacology (Berl) 109, 121-126.

41. Silverman, K., Mumford, G.K., and Griffiths, R.R. (1994). Enhancing caffeine reinforcement by behavioral requirements following drug ingestion. Psychopharmacology (Berl) 114, 424-432.

42. Horne, J.A., and Reyner, L.A. (1996). Counteracting driver sleepiness: effects of napping, caffeine, and placebo. Psychophysiology 33, 306-309.

43. Strain, E.C., Mumford, G.K., Silverman, K., and Griffiths, R.R. (1994). Caffeine dependence syndrome. Evidence from case histories and experimental evaluations. JAMA 272, 1043-1048.

44. Palatini, P., Ceolotto, G., Ragazzo, F., Dorigatti, F., Saladini, F., Papparella, I., Mos, L., Zanata, G., and Santonastaso, M. (2009). CYP1A2 genotype modifies the association between coffee intake and the risk of hypertension. J Hypertens 27, 1594-1601.

45. Palatini, P., Dorigatti, F., Santonastaso, M., Cozzio, S., Biasion, T., Garavelli, G., Pessina, A.C., and Mos, L. (2007). Association between coffee consumption and risk of hypertension. Ann Med 39, 545-553.

46. Manolio, T.A., Brooks, L.D., and Collins, F.S. (2008). A HapMap harvest of insights into the genetics of common disease. J Clin Invest 118, 1590-1605.

47. Zeggini, E., Scott, L.J., Saxena, R., Voight, B.F., Marchini, J.L., Hu, T., de Bakker, P.I., Abecasis, G.R., Almgren, P., Andersen, G., et al. (2008). Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet 40, 638-645.

48. Aulchenko, Y.S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I.M., Pramstaller, P.P., Penninx, B.W., Janssens, A.C., Wilson, J.F., Spector, T., et al. (2009). Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat Genet 41, 47-55.

49. Aulchenko, Y.S., Struchalin, M.V., Belonogova, N.M., Axenovich, T.I., Weedon, M.N., Hofman, A., Uitterlinden, A.G., Kayser, M., Oostra, B.A., van Duijn, C.M., et al. (2009). Predicting human height by Victorian and genomic methods. Eur J Hum Genet 17, 1070-1075.

50. Isaacs, A., Sayed-Tabatabaei, F.A., Aulchenko, Y.S., Zillikens, M.C., Sijbrands, E.J., Schut, A.F., Rutten, W.P., Pols, H.A., Witteman, J.C., Oostra, B.A., et al. (2007). Heritabilities, apolipoprotein E, and effects of inbreeding on plasma lipids in a genetically isolated population: the Erasmus Rucphen Family Study. Eur J Epidemiol 22, 99-105.

51. Weedon, M.N., and Frayling, T.M. (2008). Reaching new heights: insights into the genetics of human stature. Trends Genet 24, 595-603.

52. Hu, N., Wang, C., Hu, Y., Yang, H.H., Giffen, C., Tang, Z.Z., Han, X.Y., Goldstein, A.M., Emmert-Buck, M.R., Buetow, K.H., et al. (2005). Genome-wide association study in esophageal cancer using GeneChip mapping 10K array. Cancer Res 65, 2542-2546.

53. Thorleifsson, G., Walters, G.B., Gudbjartsson, D.F., Steinthorsdottir, V., Sulem, P., Helgadottir, A., Styrkarsdottir, U., Gretarsdottir, S., Thorlacius, S., Jonsdottir, I., et al. (2009). Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nat Genet 41, 18-24.

54. Dawber, T.R., Meadors, G.F., and Moore, F.E., Jr. (1951). Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health Nations Health 41, 279-281.

55. Hofman, A., Breteler, M.M., van Duijn, C.M., Krestin, G.P., Pols, H.A., Stricker, B.H., Tiemeier, H., Uitterlinden, A.G., Vingerling, J.R., and Witteman, J.C. (2007). The Rotterdam Study: objectives and design update. Eur J Epidemiol 22, 819-829.

56. Psaty, B.M., O'Donnell, C.J., Gudnason, V., Lunetta, K.L., Folsom, A.R., Rotter, J.I., Uitterlinden, A.G., Tamara B. Harris, Whitteman, J.C.M., and Boerwinkle, E. (2009). Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium Design of Prospective Meta-Analyses of Genome-Wide Association Studies From 5 Cohorts Circ Cardiovasc Genet 2, 73-80.

57. Stolk, R.P., Rosmalen, J.G., Postma, D.S., de Boer, R.A., Navis, G., Slaets, J.P., Ormel, J., and Wolffenbuttel, B.H. (2008). Universal risk factors for multifactorial diseases: LifeLines: a three-generation population-based study. Eur J Epidemiol 23, 67-74.

58. International Schizophrenia, C., Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., and Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460, 748-752.

59. Miettinen, O.S. (2009). Up from 'false positives' in genetic-and other-epidemiology. Eur J Epidemiol 24, 1-5.

60. Evans, D.M., Visscher, P.M., and Wray, N.R. (2009). Harnessing the Information Contained Within Genome-wide Association Studies to Improve Individual Prediction of Complex Disease Risk. Hum Mol Genet.

61. Janssens, A.C. (2008). Is the time right for translation research in genomics? Eur J Epidemiol 23, 707-710.

62. Drame, M., Novella, J.L., Lang, P.O., Somme, D., Jovenin, N., Laniece, I., Couturier, P., Heitz, D., Gauvain, J.B., Voisin, T., et al. (2008). Derivation and validation of a mortality-risk index from a cohort of frail elderly patients hospitalised in medical wards via emergencies: the SAFES study. Eur J Epidemiol 23, 783-791.

63. Janssens, A.C., Aulchenko, Y.S., Elefante, S., Borsboom, G.J., Steyerberg, E.W., and van Duijn, C.M. (2006). Predictive testing for complex diseases using multiple genes: fact or fiction? Genet Med 8, 395-400.

64. van Hoek, M., Dehghan, A., Witteman, J.C., van Duijn, C.M., Uitterlinden, A.G., Oostra, B.A., Hofman, A., Sijbrands, E.J., and Janssens, A.C. (2008). Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study. Diabetes 57, 3122-3128.

65. Demirkan, A., Penninx, B.W., Hek, K., Wray, N.R., Amin, N., Aulchenko, Y.S., van Dyck, R., de Geus, E.J., Hofman, A., Uitterlinden, A.G., et al. (2010). Genetic risk profiles for depression and anxiety in adult and elderly cohorts. Mol Psychiatry.

66. Ramdas, W.D., Amin, N., van Koolwijk, L.M., Janssens, A.C., Demirkan, A., de Jong, P.T., Aulchenko, Y.S., Wolfs, R.C., Hofman, A., Rivadeneira, F., et al. (2010). Genetic architecture of open angle glaucoma and related determinants. J Med Genet.

67. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42, 565-569.

68. Goldstein, D.B. (2009). Common genetic variation and human traits. N Engl J Med 360, 1696-1698.

69. Cirulli, E.T., and Goldstein, D.B. (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet 11, 415-425.

70. Kong, A., Steinthorsdottir, V., Masson, G., Thorleifsson, G., Sulem, P., Besenbacher, S., Jonasdottir, A., Sigurdsson, A., Kristinsson, K.T., Frigge, M.L., et al. (2009). Parental origin of sequence variants associated with complex diseases. Nature 462, 868-874.

71. Kong, A., Masson, G., Frigge, M.L., Gylfason, A., Zusmanovich, P., Thorleifsson, G., Olason, P.I., Ingason, A., Steinberg, S., Rafnar, T., et al. (2008). Detection of sharing by descent, long-range phasing and haplotype imputation. Nat Genet 40, 1068-1075.

72. Siva, N. (2008). 1000 Genomes project. Nat Biotechnol 26, 256.

73. Sobreira, N.L., Cirulli, E.T., Avramopoulos, D., Wohler, E., Oswald, G.L., Stevens, E.L., Ge, D., Shianna, K.V., Smith, J.P., Maia, J.M., et al. (2010). Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene. PLoS Genet 6, e1000991.

74. Walsh, T., Shahin, H., Elkan-Miller, T., Lee, M.K., Thornton, A.M., Roeb, W., Abu Rayyan, A., Loulus, S., Avraham, K.B., King, M.C., et al. (2010). Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82. Am J Hum Genet 87, 90-94.

Chapter 6

Samenvatting van dit proefschrift

Gedrag betreft het handelen van een organisme, meestal in relatie to de omgeving. Bij mensen lijkt gedrag primair bepaald te worden door het endocriene stelsel en het zenuwstelsel. De complexiteit van gedrag bij mensen is gecorreleerd aan de complexiteit van het zenuwstelsel. Gedrag kan aangeboren zijn, of aangeleerd. Veel vormen van gedrag zijn in hoge mate erfelijk, zowel binnen een soort als tussen soorten onderling.

In dit proefschrift heb ik een poging gedaan om de genetica te doorgronden van eigenschappen met een sterk gedragsaspect en waarvan bekend is dat ze de sociale, somatische, en psychologische gezondheid beïnvloeden. Ik heb een breed aantal eigenschappen bestudeerd, waaronder persoonlijkheid, *attention deficit hyperactivity disorder* (ADHD), koffie drinken en slaap. Deze zijn alle algemeen voorkomende kwantitatieve eigenschappen en het is bekend ze in belangrijke mate erfelijk bepaald zijn. Ik heb gebruik gemaakt van verschillende technieken voor het in kaart brengen van genen waaronder *linkage*- en associatieanalyse. Daarnaast heb ik gebruik gemaakt van geavanceerdere moleculaire and statistische analyses, waaronder *copy number variation* (CNV) en analyse van genoom *imprinting* voor het zoeken van genetische varianten die geassocieerd zijn met deze eigenschappen. Met deze methoden heb ik nieuwe loci (genen) ontdekt welke betrokken zijn bij deze eigenschappen.

De eerste eigenschap die in dit proefschrift is onderzocht is ADHD. ADHD is een psychiatrische ontwikkelingsstoornis met een zeer sterke gedragscomponent, resulterend in een verminderd sociaal en/of academisch functioneren. In **hoofdstuk 2** rapporteren wij resultaten van een genoomwijde *linkage*-analyse van ADHD in negen kinderen van verwante ouders uit het programma voor Genetisch Onderzoek in gelsoleerde Populaties (GRIP). Gebruikmakend van het in kaart brengen van homozygoziteit vonden we suggestief bewijs voor *linkage* (*log of odds* (*LOD*) = 2.2) van ADHD met een nieuwe locus op 18q21-22. Deze uitkomst werd verder ondersteund door de genoomwijde associatieanalyse uitgevoerd door het *Genetic Association Information Network* (GAIN) dat een verband vond met een *Single Nucleotide Polymorphism* (SNP), een variatie in één basepaar, in deze regio (rs2311120, p-value = $1.06*10^{-05}$). In een subanalyse van kinderen met ADHD van het onoplettende type vonden we ook suggestief bewijs van *linkage* met 6p22 (*LOD* = 2), een gebied dat eerder gevonden was voor ADHD bij patiënten met een leeshandicap. Andere gebieden met *HLOD* ≥ 1 zijn 1p36, 5q33, 6p12, 18p11 and 15q25. Hoewel niet significant, overlappen de meeste van deze gebieden, waaronder 1p36, 5q33 and 6p12, de gebieden die in eerdere studies werden gevonden, of liggen ze er vlak naast. Deze gebieden bevatten serotonine receptoren waaronder *HTR1B*, *HTR1E*, *HTR4*, *HTR1D* en *HTR6*. Dit wijst erop dat *linkage*-analyse een krachtige aanpak is voor de identificatie van genetische loci aangezien we verschillende interessante kandidaat-gebieden konden ontdekken met slechts negen patiënten vergeleken met de 958 trio's in de GAIN analyse.

Vervolgens heb ik mij gefocust op persoonlijkheid. De structuur van de menselijke persoonlijkheid wordt traditioneel toegeschreven aan een relatief kleine groep kenmerken. Gedurende de afgelopen eeuw is de wetenschappelijke consensus toegegroeid naar een

taxonomisch model van persoonlijkheidskenmerken gebaseerd op vijf hogere orde dimensies als neuroticisme, extraversie, het openstaan voor ervaringen, plezierigheid en zorgvuldigheid, bekend als het Vijf Factoren Model (*Five Factor Model, FFM*). Deze vijf dimensies zijn ontwikkeld als orthogonale schalen en daardoor grotendeels onafhankelijk. Gezamenlijk leveren ze een brede beschrijving van persoonlijkheid. Neuroticisme wordt gewoonlijk gedefinieerd als emotionele instabiliteit; het betreft het ervaren van negatieve emoties zoals overmatige bezorgdheid, depressie, vijandigheid en gevoeligheid voor stress. Extraversie wordt gekarakteriseerd door positieve emoties, de neiging om actief te zijn, het zoeken naar stimulans en het genieten van het gezelschap van anderen. Openstaan voor ervaringen omvat een actieve verbeelding, aandacht voor esthetiek, een voorkeur voor variatie en intellectuele nieuwsgierigheid. Plezierigheid kan worden gedefinieerd als de neiging tot samenwerken en medelevend zijn in plaats van wantrouwend en antagonistisch jegens anderen. De dimensie zorgvuldigheid, tenslotte, betreft zelfdiscipline, nauwgezetheid, grondigheid, gestructureerdheid, bedachtzaamheid en het verwezenlijken van zaken. Persoonlijkheidskenmerken voorspellen tal van aan uitkomsten op sociaal, gedrags- en gezondheidsgebied, zoals prestaties op het werk, hoge ouderdom en vele psychiatrische aandoeningen, waaronder misbruik en afhankelijkheid van middelen, stemmingsstoornissen zoals depressie, angststoornissen en persoonlijkheidsstoornisen. Het vinden van genen waaraan persoonlijkheid kan worden toegeschreven is moeilijk gebleken ondanks de grote erfelijke component, waarbij de schattingen voor de erfelijkheid uiteenlopen van 33 % tot 65 %.

Gedurende mijn promotieonderzoek heb ik meegewerkt aan een grote genoomwijde associatiestudie (GWA) met als doel het identificeren van veelvoorkomende genetische variaties met kleine effecten. In de analyses werden tien groepen voor eerste analyse (17,375 volwassenen) en vijf *in-silico* replicatiegroepen voor het bevestigen van de bevindingen (3,294 volwassenen) meegenomen. Alle participanten waren van Europese afkomst. Persoonlijkheidsscores voor neuroticisme, extraversie, openheid voor ervaringen, plezierigheid en zorgvuldigheid waren gebaseerd op de NEO vijf-factoren schaal (*Neo Five-Factor Inventory*, NEO-FFI). Genotype-data was beschikbaar van ca. 2.4 miljoen SNPs (zowel direct getypeerd als geïmputeerd met behulp van HAPMAP data). In de groepen voor eerste analyse werden klassieke associatieanalyses uitgevoerd op basis van een additief model, gevolgd door meta-analyse gebruikmakend van de gewogen inverse-variantiemethode. De resultaten lieten genoomwijde significantie zijn voor het openstaan voor ervaringen in de buurt van het *RAS1* gen op 5q14.3 (rs1477268 en rs2032794, p-value = $2.8*10^{-08}$ en p-value = $3.1*10^{-08}$) en voor zorgvuldigheid in het *KATNAL2* gen op 18q21.1 (rs2576037, p-value = $4.9*10^{-08}$) dat tot uitdrukking komt in de hersenen. Daarnaast hebben we een gen-gebaseerde test uitgevoerd welke de associatie bevestigde tussen *KATNAL2* en zorgvuldigheid. *In silico* replicatie liet echter geen significante associaties zien van de top SNPs met openheid en zorgvuldigheid, hoewel de richting van het effect van de *KATNAL2* SNP op zorgvuldigheid wel consistent was in alle replicatiesamples. GWA studies op grotere schaal

en een alternatieve aanpak zijn nodig om te bevestigen dat *KATNAL2* een nieuw gen is dat betrekking heeft heeft op zorgvuldigheid.

Voor dit proefschrift heb ik een alternatieve aanpak gevolgd. Om genetische loci te vinden welke mogelijk middelgrote tot grote effecten hebben op de persoonlijkheidskenmerken neuroticisme, extraversie, openheid, plezierigheid en zorgvuldigheid, heb ik in **hoofdstuk 3.1** een genoomwijde *linkage*-scan uitgevoerd van hoog-scorende individuen (percentiel) in de Erasmus Rucphen Familie studie (ERF) voor elk van de vijf kenmerken. Wij vonden significant bewijs voor *linkage* van zorgvuldigheid met 20p13 bij een recessief model (LOD = 5.86). Haplotype-constructie leverde een specifiek haplotype op dat overerfde in zes families met hoge LOD score-bijdragen, welke een marginale associatie vertoonden (p-value = 0.058 en p-value = 0.038) bij gebruik van een recessief model. Dit haplotype beslaat een gebied van 100 kilobase (kb) op 20p13 en bevat vijf genen behorend tot de familie van beta defensinen (*DEFB*). Wij vonden ook suggestief bewijs van *linkage* van neuroticisme met 21q22 (LOD = 3.42, alwaar we twee significante haplotypen vonden (met respectievelijk p-value = 0.009 en p-value = 0.007). De hoogst scorende individuen uit drie families met een hoge LOD-score bijdrage hadden een homozygotie voor één van deze haplotypen gemeen. Dit gebied bevat tien genen waaronder *TTC3*, *DSCR9*, *DSCR3*, *DYRK1A*, *KCNJ6* en *PIGP*, en wordt geassocieerd met cognitieve en mentale achteruitgang in individuen met het syndroom van Down. Voor plezierigheid vonden wij suggestieve *linkage* (*LOD* = 3.32), maar wel een significant haplotype (p-value = 0.018) in het gebied 17q24. Genen in dit gebied zijn ondermeer *KCNJ2*, *KCNJ16* en *MAP2K6*. Overmatige homozygositeit in elk van deze gevonden gebieden deed ons zoeken naar deleties in deze regios. Echter, er bleek uit een analyse van variaties in het aantal kopieën (*copy number variation*, CNV) geen bewijs te zijn voor CNVs. Deze gebieden kwamen niet naar voren in de eerder genoemde GWAS. Hiervoor kunnen verschillende verklaringen voor zijn, waaronder een fout-positief resultaat in deze studie. Het is echter meer voor de hand liggend dat GWA studies vermoedelijk niet in staat zijn om relatief zeldzame recessieve varianten op te pikken. Zulke varianten zouden daadwerkelijk een deel van de onverklaarde erfelijkheid (*missing heritability*) kunnen verklaren in kwantitatieve eigenschappen.

In **hoofdstuk 3.2** hebben we een alternatieve *linkage*-analyse uitgevoerd waar in de verschillende persoonlijkheidsuitkomsten kwantitatief werden geanalyseerd. Hoewel analyse van kwantitatieve eigenschappen een krachtige aanpak is, is het aantal te onderzoeken individuen van groot belang. Bovendien is replicatie ook voor *linkage*-analyse van belang om fout-positieve resultaten uit te sluiten. Daarom hebben we een meta-analyse van vier niet eerder gepubliceerde *linkage*-analyses van kwantitatieve eigenschappen van de vijf persoonlijkheden uitgevoerd. De samples (*N* = 6,149) bestonden uit de Erasmus Ruchpen Familie studie (ERF), het Nederlands Tweelingen Register (NTR) en een sample met volwassenen en adolescenten uit Australië (QIMR). Voor een zo hoog mogelijke power hebben we methoden voor genoombrede *linkage*- en associatieanalyse gecombineerd in een tweestaps analyse. De

gebieden met significant en suggestief bewijs voor *linkage* werden fijner in kaart gebracht met behulp van de resultaten van de grootste meta-analyse van genoombrede associatiestudies (*N* > 18,000) van deze vijf persoonlijkheidskenmerken. Significante *linkage* werd gevonden voor neuroticisme op chromosoom 3p14 (*LOD* = 4.67) en op chromosoom 19q13 (*LOD* = 3.55), voor extraversie op 14q32 (*LOD* = 3.3); voor plezierigheid op 3p25 (*LOD* = 3.67) en op twee naast elkaar gelegen gebieden op chromosoom 15, waaronder 15q13 (*LOD* = 4.07) and 15q14 (*LOD* = 3.52) in de individuele scans. In de meta-analyse vonden we sterk bewijs voor *linkage* van extraversie met 4q34, 9q34, 10q24 and 11q22, openheid met 2p25, 3q26, 9p21, 11q24, 15q26 and 19q13, en plezierigheid met 4q34 and 19p13. Na combineren van deze gegevens met de associatieresultaten van de GWAS van deze persoonlijkheidskenmerken detecteerden we voldoende bewijs voor openheid met 11q24 (rs677035, p-value = $2.6*10^{-06}$). We vonden bewijs met bijna statistisch significante voor associatie tussen neuroticisme en rs332389 (p-value = $9*10^{-05}$) op 3p14, en tussen extraversie en rs7088779 (p-value = $4.2*10^{-06}$) op 10q24. Van deze drie SNPs is rs677035 een tussen twee genen gelegen SNP tussen *FLI1* en *KCNJ1*, rs332389 ligt in het intron van het gen *SLC25A26* en rs7088779 is gesitueerd tussen *CRTAC1* (*cartilage acidic protein* 1) en *C100rf28*, een regio welke eerder geassocieerd is met de ziekte van Alzheimer. Ook in de kwantitatieve *linkage*-analyse, welke meer gelijk is aan GWAS, was er geen bewijs voor een rol van *KATNAL2* or *RASA1* voor zorgvuldigheid en openheid. Wederom kan er gesteld worden dat dit verklaard kan worden door een fout-positief resultaat of door verschillen in power voor de diverse varianten betrokken bij deze complexe eigenschappen. Een interessante familie van genen die tevoorschijn komt uit de *linkage*-studies in dit proefschrift is de KCNJ familie. De door ons, door middel van analyse van dichotome en kwantitatieve uitkomsten gevonden *linkage*-gebieden bevatten opvallend veel KCNJ genen, waaronder *KCNJ6*, *KCNJ16*, *KCNJ2* and *KCNJ1*. KCNJs zijn kalium kanalen die de potentiaal naar binnen toe herstellen. Ze zijn doelen van meerdere toxines en het disfunctioneren van deze kanalen wordt in verband gebracht met verschillende ziekten waaronder neurologische en endocriene aandoeningen en aritmieën op het elektrocardiogram.

Epigenetische effecten, waarvan *imprinting* van het genoom een van de vormen is, wordt vaak aangehaald als reden voor verlies aan efficiëntie van genoomwijde associatie studies voor het identificeren van veelvoorkomende varianten geassocieerd met complexe eigenschappen. In **hoofdstuk 3.3** ondernamen we een genoomwijde zoektocht naar gebieden met *imprinting* voor persoonlijkheidskenmerken in het ERF sample gebruikmakend van SNPs van het Illumina 6K *linkage panel*. We vonden significante verschillen in paternaal en maternaal doorgegeven allelen op 6q25 (rs602890, p-value = $5*10^{-07}$) voor extraversie en op 4q21 (rs1880719, p-value = $8*10^{-07}$) voor plezierigheid. De twee SNPs, welke niet significant waren in een algemene associatietest, vertoonden wel significante associatie indien de parentale herkomst van de allelen meegenomen werd in het associatiemodel (p-value = $3*10^{-06}$ voor rs602890 en p-value =$1*10^{-06}$ voor rs1880710). De SNP rs602890 vertoonde tegenovergestelde en nominaal

significante effecten indien geërfd van vader (β = -2.78, p-value = $2.7*10^{-03}$) en moeder (β = 4.33, p-value = $2*10^{-04}$) en rs1880719 vertoonde alleen significante associatie met plezierigheid indien geërfd van de moeder (β = -4.47, p-value = $6*10^{-05}$). Het 6q25 gebied staat bekend als een gebied met *imprinting*, terwijl *imprinting* ook verondersteld wordt als reden voor de te sterke groei van het centrale zenuwstelsel in het 4q21/4q23 syndroom. SNP rs602890 ligt in het intron van het *ZDHHC14* gen en rs1880719 is een SNP in het intron van het *SCD5* gen dat codeert voor stearoyl-CoA desastrase. Geen van deze SNPs vertoonde enig bewijs van associatie in de grote GWAS of *linkage*-studies als ouderschapseffecten buiten beschouwing werden gelaten. Deze resultaten kunnen slechts gerepliceerd worden in een familiegebaseerd sample. Een poging om de twee interessantste resultaten (rs1880719 and rs602890) te repliceren in de Australische tweeling samples leverde onvoldoende bewijs op voor ouderschapseffecten. We nemen aan dat dit komt door de ver uiteen liggende markers die we in onze analyse gebruikten (*N* = 5,250), hetgeen het moeilijk maakte om het gebied precies aan te wijzen. Het op fijnere schaal in kaart brengen van het 4q21 gebied resulteerde in additionele significante SNPs (in hoog LD met rs1880719) in hetzelfde gen (*SCD5*). Momenteel wordt er gewerkt aan een andere replicatie in het tweelingen sample van het Nederlands Tweelingen Register (NTR).

Tenslotte hebben we in dit proefschrift slaap en cafeïne bestudeerd. Epidemiologische studies hebben aangetoond dat slaapduur geassocieerd is met verminderd psychosomatisch functioneren. In **hoofdstuk 4.2** hebben we de eerste grootschalige genoomwijde associatieanalyse uitgevoerd voor slaapduur in 7 Europese cohorten (*N* = 4,260). We hebben een variant in een intron geïdentificeerd op chromosoom 21p12.1 (rs11046205) in het *ABCC9* gen en we hebben deze variant gevalideerd in een replicatiestudie. Zodra de geconserveerde *Drosophila* homoloog werd uitgeschakeld pan-neuronisch door RNAi, werd de nachtslaapduur gereduceerd. *ABCC9* codeert voor een subeenheid van een ATP-gevoelig kalium kanaal, welke functioneert als sensor van intracellulaire energie. Op die manier verbinden onze resultaten slaap en haar herstellende effecten, in fysiologische termen bekend als de homeostaat, met energiemetabolisme in de cel. De tijd tot het in slaap vallen (slaaplatentietijd) is een belangrijke determinant voor de kwaliteit van slaap. Toename van slaaplatentietijd is een van de voornaamste kenmerken van *sleep onset insomnia* en *delayed sleep phase syndrome*, twee veelvoorkomende circadiane stoornissen welke het sociale en academische leven en de arbeidsprestaties van een individu hinderen. Waar eerder nog geen genetische onderzoeken gedaan waren naar deze eigenschap, hebben wij in **hoofdstuk 4.1** een meta-analyse uitgevoerd van zeven genoomwijde associatiestudies bestaande uit 4,270 individuen van Europese origine. We vonden sterk bewijs voor de associatie van een variant (rs9900428, p-value = $5.7*10^{-08}$) van het hexaribonucleotide bindend proteïne 3 gen (*HRNBP3*, ook bekend als *FOX3*) met slaaplatentietijd. We hebben onze resultaten succesvol gerepliceerd in vijf additionele cohorten (*N* = 10,901). De meta-analyse van het eerste-analyse cohort en de replicatiecohorten toonde een genoomwijde significante associatie van rs9900428 (p-value = $3.8*10^{-08}$). Het *FOX3* gen codeert voor een sterk hersen-

specifiek eiwit en behoort tot de familie van de FOX genen, welke regulatoren zijn van pre- en postnatale alternative eiwitsplitsingen in spier- en hersenweefsel. Het *FOX3* eiwit wordt verondersteld een stroomopwaartse regulator te zijn voor alternatieve splitsing van hersen-specifieke mRNAs die betrokken zijn in neurotransmissie via glutamaten en GABA. Dit is de eerste genoomwijdezoektocht naar slaaplatentietijd en er was eerder nog geen variant voor slaaplatentietijd gerapporteerd.

We hebben tenslotte een belangrijke determinant van slaaplatentietijd bestudeerd, cafeïne. Van cafeïne is bekend dat het effect heeft op gedrag indien ingenomen in matige hoeveelheden. Hoge doses cafeïne kunnen ook negatieve effecten hebben zoals nervositeit, rusteloosheid en slapeloosheid. De inname van cafeïne wordt ook gebruikt als model voor drugsmisbruik, en koffie is de belangrijkste bron van cafeïne in volwassenen. In **hoofdstuk 4.3** hebben we een grootschalige meta-analyse gedaan van genoomwijde associatiestudies van koffie inname. Onze data (N > 18,000) bestond uit samples van tien cohorten. We vonden genoomwijd significant bewijs voor associatie van een variant (rs2470893) in het 15q24 gebied, gelegen tussen de genen *CYP1A1* en *CYP1A2*. Het "T" allel van rs2470893 werd positief geassocieerd met koffie drinken. Het *CYP1A1* gen werd ook minder actief na cafeïnebehandeling in lymphoblastoïde cellijnen. Van *CYP1A1* is bekend dat het polycyclische aromatische koolwaterstoffen zoals benzo(a)pyreen metaboliseert, hetgeen een belangrijk bestanddeel van koffie is en waarvan ook bekend is dat het betrokken is bij de mutagenese als gevolg van koffie inname, terwijl *CP1A2* betrokken is bij de primaire afbraak van cafeïne. Gecombineerd bewijs van GWAS en cafeïne expressie analyse suggereert dat *CYP1A1* het drinken van koffie reguleert en dat deze associatie onafhankelijk is van rookstatus. Naast *CYP1A1/CYP1A2* vonden we ook sterk bewijs voor associatie van koffie-inname met varianten in de genen *NCALD* (rs16868941), *NRCAM* (rs382140) en *CAB39L* (rs9526558). Eerder vonden we een sterke associatie van het *NCALD* gen met slaaplatentie (rs17498920, p-value = $2*10^{-05}$). Dit is een interessante bevinding omdat het bekend is dat koffie-inname interfereert met melatoninesecretie en het begin van de slaap vertraagt. Een replicatie analyse van de top bevindingen van de GWAS in 8,000 individuen van Nederlandse afkomst (LifeLines), bevestigde associatie van het *CYP1A1/A2* locus en rs382140 in de buurt van het *NRCAM* gen met het drinken van koffie. *CAB39L*, hoewel niet succesvol gerepliceerd, bleek te zijn opgereguleerd in alle drie de cellijnen die bestudeerd waren na cafeïne behandeling wat suggereert dat het gen inderdaad betrokken is bij cafeïne metabolisme en –inname. Associatieanalyse van de varianten in deze genen met de koffie gerelateerde fenotypen bloeddruk, mortaliteit, kankermortaliteit, lipiden en de ziekte van Alzheimer liet significante associatie van rs2470893 (*CYP1A1/1A2*) met systolische en diastolische bloeddruk zien waar wederom het "T" allel positief geassocieerd was, wat suggereert dat verhoogde koffie-inname de bloeddruk verhoogt. De associatie van onze top hits met bloeddruk is zeer interessant en versterkt het bewijs dat koffieconsumptie de bloeddruk verhoogt.

Chapter 7

Acknowledgements

As I write this part of my thesis I can't help diving back five years when I came to the Netherlands to start my research work at the Department of Epidemiology. With only a master's degree in Statistics in hand and almost non existing knowledge of genetics, biology or diseases, I started my research work at the Genetic Epidemiology Unit of the Department of Epidemiology. These 5 years have been full of knowledge, adventure and excitement yet sometimes tiring. This thesis, knowledge, adventure and excitement would not have been possible without the help of many people to whom I would like to express my gratitude in person. But as a start, I take this opportunity to acknowledge the contribution of all the individuals who supported me during this period.

Above all, my promotor Professor Cornelia M. van Duijn. Dear Cornelia, you have been more than just a promotor for me. Not only that you are a great scientist, you are also a great teacher and a great role model, in fact you are my guru. Despite your extremely busy schedules you always have time for your students and you can spend your time and resources trying to solve the problems of your students whether academic or personal. It is your personality, knowledge and enthusiasm that makes this unit a wonderful place to work. And this is what makes me enjoy every minute of my work here. The confidence that you put in me is the probably the sole reason of my growth as a scientist. It has been an honour and a great pleasure and a lot of fun to work for you. Thank you for giving me the opportunity to be a part of your group.

My sincere thanks to my second promotor, Professor Ben A. Oostra. Dear Ben your support has been crucial, especially during the period when I was finishing this thesis.
Your great organization and management skills that you keep teaching us from time to time, your quick and timely responses and guidance helped me complete this thesis in time without getting tired or stressed. Your directions have always been very clear, precise and helpful. Thank you for everything that you did for me.

I would like to offer special thanks to Professor Dorret I. Boomsma, Professor Till Roenneberg and Dr. Henning Tiemeier for being a part of doctoral committee. Dear Dorret, it is an honour for me to work with you on several projects including personality and ADHD. Dear Till, it was a pleasure working with you on the sleep projects and I look forward to work with you more on such projects. Dear Henning, thank you for all the time you invested in reading my thesis, it was a great pleasure to work with you on sleep and coffee projects. I also feel honoured to have Professor Jurgen Del-Favero in the large committee. Thank you for investing your time an effort in this promotion. A note of thanks to Dr. Cecile Janssens, for being a part of my large committee, and for all the intellectual suggestions that I received from you on my papers. I really appreciate your comments, which always made my (ridiculously long) papers comprehensible. I also enjoyed those non-scientific chats blended with your lovely sense of humour.

Professor Albert Hofman, it was a privilege and an honour to be taught by you. Dear Yurii, being my supervisor for my first paper, you taught me how to write scientific papers and how to analyze the data. I still remember that you literally had to force me to learn 'R' and I am glad that you did. Thank you for all your intellectual output, suggestions and discussion full of knowledge. It was a pleasure and a great learning experience to assist you in GE02 and GE03 courses, it taught me something new every time. Thank you also for giving me an opportunity to teach the GE02 course.

A special thanks to Jeannette Lokker for all the paper work and communication that you did concerning this promotion, which helped me concentrate on my thesis. Without you I would have been lost.

My dear paranimf and friend Maaike Schuur. Dear Maaike it was lovely to share office with you for 2 years. It was the best period of my PhD here. In fact it was you who made me feel at home in the Netherlands. I really enjoyed our coffee time and our regular monthly dinners. I am glad that despite your busy schedules in the clinic, we still see each other regularly and get some time for dinners; I hope that we'll continue with these traditions. Thanks for being here.

My dear paranimf, friend and coffee buddy, Lennart Karssen. Dear Lennart, I could not ask for a better room mate than you after Maaike left. Thank you for all the support that you gave me when I was finishing this thesis. This support came at a time when I needed it the most and shows that you are truly a friend in need (a friend indeed!). Besides, your great sense of humour kept me cheerful all along. Thank you for the Dutch translation of the summary of this thesis and also for the beautiful scripts that you wrote for me, and thanks for being my paranimf.

Dear Aaron, you have been a great buddy. As I recall, there was hardly a single time when I asked you something and you did not help. You have always been there and were always willing to help. I enjoyed every bit of our scientific and non-scientific discussions and chats (especially the Friday ones!). Thanks for the R scripts that you wrote for me.

Dear Ayse, in you I have found a good friend. Thank you for all the work that you did for the sleep paper and also for taking some work off my shoulders when I was finishing. I really enjoy our frequent dinning out and hope that we will continue doing that in the future. Dear Maksim, thank you for all your help with the 'R' and Linux scripts. I really appreciate the fact that no matter how busy you were, you were always willing to solve my scripting and GenABEL problems. Dear Linda, you are a nice person; always ready to help others. Thanks for all the work that you did for me for the coffee paper. I wish you success with your PhD. Dear Sara and Carla, you guys are a nice and cheerful addition to the Genetic Epi group and I wish you all the best and success, a

special thanks to both Carla and Sara also for helping Lennart with the Dutch translation of the summary of this thesis. Dear Elza, I am glad that we have a secretary like you, so enthusiastic and so organized. Thanks for all your work and support.

I would also like to thank my colleagues from the laboratory. Dear Jeanette Vergeer you are the most crucial person in this group; a person who knows exactly where to find the stuff that we are looking for (data). Without you, I think we all will be lost. Dear Andy, Andrea Vermij, Petra Veraart and Bernadette, thank you all for the work, which was crucial for the work in this thesis. Dear Dick Slof, thank you for the logistic support.

I would like to thank our Russian collaborators. Dear Professor Tatiana Axenovich, it has been a pleasure and a great learning experience to work with you on linkage projects. There is a lot from you still that I would like to learn. Dear Nadezhda Belonogova, Irina Zorkoltseva and Anatoly Kirichenko, thank you for the work that you did for my papers and I look forward to working with you more.

It has been a great pleasure to work with my part time colleagues Leonieke Koolwijk, Hilde Peeters, Peter Henneman, Boukje de Vries and Hille Fieten and Natalia Rivera. Also my colleagues from the 21st floor, Wishal Ramdas, Karin Hek, Germaine Verwoert, Stefan Walter, Raluca Mihaescu, Abbas Dehghan, Arfan Ikram, it has been a pleasure to work with you guys.

I am greatly indebted to my ex-colleagues with whom I had my first hand learning experience. Dear Fan, sharing the room with you for about 2 years I learned many things about genetic epidemiology from you. You have been like a teacher to me, a helping colleague and a good friend. I enjoyed the dinners with you and Bei and the nice time that we spent at the Scheveningen beach together. Alejandro, I remember that you gave me a one and a half hour crash course about linkage analysis that was really helpful. It was a pleasure to work with my colleagues Fakhredin Syed, Sandra Lopez, Angela Gonzalez, Luba Pardo, Mojgan Yazdanpanah, Slavica Peckioska, Dominiek Despreit, Wing Choy, Annelous Berends, Mark Sie, Suzanne Gelok, Helena Schmidt, Tonnie Coppus, Sophie van Wingerden, Moniek van Zitteren, Pratika Hernanda, Pascual Sanchez, Elena Gusareva, Nicola, Martin Ladouceur, Debby, Hasna, Sue Ellen and the secretaries Annette Bergman-Tolhoek and Brigitte van der Knaap.

My sincere thanks to Bert Eussen and Annelies de Klein who taught me how to perform copy number variation analysis using array data. It was also a great pleasure to work with Dr. Carola Zillikens and Dr. Fernando Rivadeneira on GIANT and GEFOS projects.

I would like to acknowledge the secretarial and cluster bureau 13 staff including Yolanda Bekker, Marion Boltjes, Jacqueline Arnell, Hetty Gerritse-Kattouw, Erica Kroos, Annette Moerland Vink and Solange Degenaar for all the administrative and paper work that you guys did for me.

Dear Danielle, Elisabeth and Martje, it was really fun to share EE-2224 with you guys. Dear Danielle I really wish you success in your endeavor to win a Nobel Prize. Please do not forget me when you get that. Dear Martje, I would love to have coffee and dinner with you again (also some karate lessons). I wish you success with your internships at the hospital and in your career.

I am really indebted to the Government of Pakistan that provided me opportunity and financial support to do research in the Netherlands. Special thanks to Dr. Atta-ur-Rehman for initiating the overseas PhD projects. Also thanks to Professor Javed Shabbir and Professor Attaullah Kalim for facilitating my admission to the NIHES programme. My sincere thanks to Mrs. Minkman Loes for coordinating between higher education commission of Pakistan and the NUFFIC and facilitating my stay here.

Thanks, to our friends including Bhabi Zohra and Bhaijan Zameer for letting us stay at their house for the first 2 months when we came to the Netherlands; to Bhabi Farrukh and Bhaijan Naseer for their moral support and hospitality; to Mahrukh and Masood Bhai for their help and friendship.

Finally the two people that I love the most and who stood beside me through every thick and thin. My mother, Gul-e-Shagufta -- the iron lady. Dear ammi (mom), I don't even have the words to acknowledge what you have done for me. You have been both; a father and a mother to me, in addition to being a friend. I really wish I could give you something in return and I hope to make you proud someday. I am lucky to have a mother like you. Thank you for everything that you did for me. Last but certainly not the least my husband Amin. Dear Amin we have been in a relationship for 12 years now, and we know that these 12 years were full of struggle and hard work for both of us. I cannot forget how we used to sit with each other and make plans for our future and how we still do that. The reason why we are so enthusiastic and kicking despite so many disappointments and failures we confronted in our lives is that we have always been supportive of each other in all the ups and downs. You have been more than just a husband; you have been my best buddy, the person who makes me smile when I am down, the person who gives me strength. Thank you for being with me and supporting me in my every decision.

Chapter 8

About the Author

Najaf Amin was born in Islamabad, Pakistan on January 1st, 1979. She finished her secondary school at the Islamabad College for Girls in 1994. She continued her studies at the same college with the specialization in Mathematics, Statistics and Economics. After graduating in 1996, she completed her bachelor's degree in science in 1998 at the Punjab University Lahore. She continued her education at the Quaid-e-Azam University in 1999 and completed her Masters degree in Statistics in 2001. From 2001 to 2005 she served as a lecturer in Statistics at various institutes in Islamabad. In October 2005 she moved to the Netherlands to study genetic epidemiology at the department of Epidemiology, Erasmus University Medical Center, after getting a PhD scholarship announced by the Higher Education Commission of Pakistan. In 2007 she was conferred the degree of Doctor of Science in Genetic Epidemiology at the Netherlands Institute of Health Sciences (NIHES). She continued her research work in the department of epidemiology at Erasmus MC as a research fellow. She is planning to continue working with the department of epidemiology under the supervision of Prof. C. M. van Duijn.

PhD Portfolio Summary

Summary of PhD training and teaching activities

Name PhD student: Najaf Amin Erasmus MC Department: Epidemiology Research School: NIHES	PhD period: 2007-2011 Promotor(s): Prof. C.M. van Duijn & Prof. B.A. Oostra

1. PhD training

	Year	Workload (Hours/ECTS)
In-depth courses (e.g. Research school, Medical Training)		
Doctor of Science in Genetic Epidemiology	**2005-2007**	**30**
Erasmus Winter Programme 2006	2006	1
Erasmus Summer Programme 2007	2007	5
CC01 Study Design	2006	4.3
GE02 Genetic -Epidemiologic Research Methods	2005	5.7
GE08 SNP's and Human Disease	2005	1.8
EP15 Principles of Epidemiologic Data Analysis	2006	1.4
GE03 Advances in Population-based Studies of Complex Genetic Disorders	2006	1.4
GE04 Genetic Linkage Analysis: Model-based Analysis 2006	2006	1.4
GE05 Genetic Linkage Analysis: Model-free Analysis	2006	1.4
Presentations		
- Genetic Epidemiology Unit of Department of Epidemiology: Attention Deficit Hyperactivity Disorder(ADHD) A Genome Wide: Scan	2006	1
- Department of Epidemiology: Genome-wide linkage analysis of Attention Deficit Hyperactivity Disorder(ADHD)	2006	1
- Genetic Epidemiology Unit of Department of Epidemiology: Testing for association in families with CC-QLS	2007	1
- Genetic Epidemiology Unit of Department of Epidemiology: Potential of Genomic Control for Pedigree based quantitative trait Association Analysis	2007	1
- Department of Epidemiology: Genome-wide Association Analysis of NEO Personality Traits	2008	1
- Genetic Epidemiology Unit of Department of Epidemiology: Personality switch board: 21q???	2008	1
- Genetic Epidemiology Unit of Department of Epidemiology: Linkage Analysis of the Big Five Personality Traits	2009	1
- Genome-wide linkage analysis of personality traits	2010	1

International conferences		
- European Mathematical Genetics Meeting 2007: Potential of Genomic Control for Pedigree-Based Association of Quantitative Traits	2007	1
- European Mathematical Genetics Meeting 2008	2008	1
- 6th Annual Centre for Medical Systems Biology meeting 2009	2009	1
- Engage Meeting 2009	2009	1
- European Society of Human Genetics Conference 2010: A genome-wide linkage scan of personality traits reveals a new locus for conscientiousness	2010	1
- World Congress on Psychiatric Genetics XVIII: A genome-wide linkage scan of personality traits reveals a new locus for conscientiousness	2010	1
Seminars and workshops		
- Weekly scientific seminars of Genetic Epidemiology Unit , Department of Epidemiology	2005-2010	1
Other		
- Organisation of scientific meetings of the Genetic Epidemiology unit of the Department of Epidemiology	2007-2010	3

2. Teaching activities

	Year	**Workload (Hours/ECTS)**
- Teaching assistant of Dr. Yurii Aulchenko for GE02 course	2007-2008	30 hrs
- Teaching assistant of Dr. Yurii Aulchenko for GE03 course	2007-2008	30 hrs
- Teaching assistant of Dr. Yurii Aulchenko for the ESP29	2007-2009	30hrs
Lecturing		
- Lecturer for the GE02 course	2009-2010	60 hrs
Supervising Master's theses		
Other		
- Reviewer European Journal of Epidemiology	2009	0.5
- Reviewer Biological Psychiatry	2009	0.5
- Reviewer European Journal of Human Genetics	2009	0.5
- Reviewer American Journal of Medical Genetics	2010	0.5

.

List of publications

Amin N, Aulchenko YS, Dekker MC, Ferdinand RF, van Spreeken A, Temmink AH, Verhulst FC, Oostra BA, van Duijn CM. *Suggestive linkage of ADHD to chromosome 18q22 in a young genetically isolated Dutch population*. Eur J Hum Genet. 2009 Jul;17(7):958-66.

Amin N, van Duijn CM, Janssens AC. *Genetic scoring analysis: a way forward in genome wide association studies*? Eur J Epidemiol. 2009;24(10):585-7.

Amin N, van Duijn CM, Aulchenko YS. A genomic background based method for association analysis in related individuals. PLoS One. 2007 Dec 5;2(12):e1274.

Ramdas WD, **Amin N**, van Koolwijk LM, Janssens AC, Demirkan A, de Jong PT, Aulchenko YS, Wolfs RC, Hofman A, Rivadeneira F, Uitterlinden AG, Oostra BA, Lemij HG, Klaver CC, Vingerling JR, Jansonius NM, van Duijn CM. *Genetic architecture of open angle glaucoma and related determinants*. J Med Genet. 2010 Nov 7.

de Moor MH, Costa PT, Terracciano A, Krueger RF, de Geus EJ, Toshiko T, Penninx BW, Esko T, Madden PA, Derringer J, **Amin N**, Willemsen G, Hottenga JJ, Distel MA, Uda M, Sanna S, Spinhoven P, Hartman CA, Sullivan P, Realo A, Allik J, Heath AC, Pergadia ML, Agrawal A, Lin P, Grucza R, Nutile T, Ciullo M, Rujescu D, Giegling I, Konte B, Widen E, Cousminer DL, Eriksson JG, Palotie A, Peltonen L, Luciano M, Tenesa A, Davies G, Lopez LM, Hansell NK, Medland SE, Ferrucci L, Schlessinger D, Montgomery GW, Wright MJ, Aulchenko YS, Janssens AC, Oostra BA, Metspalu A, Abecasis GR, Deary IJ, Räikkönen K, Bierut LJ, Martin NG, van Duijn CM, Boomsma DI. *Meta-analysis of genome-wide association studies for personality. Mol Psychiatry*. 2010 Dec 21.

Elks CE, Perry JR, Sulem P, Chasman DI, Franceschini N, He C, Lunetta KL, Visser JA, Byrne EM, Cousminer DL, Gudbjartsson DF, Esko T, Feenstra B, Hottenga JJ, Koller DL, Kutalik Z, Lin P, Mangino M, Marongiu M, McArdle PF, Smith AV, Stolk L, van Wingerden SH, Zhao JH, Albrecht E, Corre T, Ingelsson E, Hayward C, Magnusson PK, Smith EN, Ulivi S, Warrington NM, Zgaga L, Alavere H, **Amin N**, Aspelund T, Bandinelli S, Barroso I, Berenson GS, Bergmann S, Blackburn H, Boerwinkle E, Buring JE, Busonero F, Campbell H, Chanock SJ, Chen W, Cornelis MC, Couper D, Coviello AD, d'Adamo P, de Faire U, de Geus EJ, Deloukas P, Döring A, Smith GD, Easton DF, Eiriksdottir G, Emilsson V, Eriksson J, Ferrucci L, Folsom AR, Foroud T, Garcia M, Gasparini P, Geller F, Gieger C; GIANT Consortium, Gudnason V, Hall P, Hankinson SE, Ferreli L, Heath AC, Hernandez DG, Hofman A, Hu FB, Illig T, Järvelin MR, Johnson AD, Karasik D, Khaw KT, Kiel DP, Kilpeläinen TO, Kolcic I, Kraft P, Launer LJ, Laven JS, Li S, Liu J, Levy D, Martin NG, McArdle WL, Melbye M, Mooser V, Murray JC, Murray SS, Nalls MA, Navarro P, Nelis M, Ness AR, Northstone K, Oostra BA,

Peacock M, Palmer LJ, Palotie A, Paré G, Parker AN, Pedersen NL, Peltonen L, Pennell CE, Pharoah P, Polasek O, Plump AS, Pouta A, Porcu E, Rafnar T, Rice JP, Ring SM, Rivadeneira F, Rudan I, Sala C, Salomaa V, Sanna S, Schlessinger D, Schork NJ, Scuteri A, Segrè AV, Shuldiner AR, Soranzo N, Sovio U, Srinivasan SR, Strachan DP, Tammesoo ML, Tikkanen E, Toniolo D, Tsui K, Tryggvadottir L, Tyrer J, Uda M, van Dam RM, van Meurs JB, Vollenweider P, Waeber G, Wareham NJ, Waterworth DM, Weedon MN, Wichmann HE, Willemsen G, Wilson JF, Wright AF, Young L, Zhai G, Zhuang WV, Bierut LJ, Boomsma DI, Boyd HA, Crisponi L, Demerath EW, van Duijn CM, Econs MJ, Harris TB, Hunter DJ, Loos RJ, Metspalu A, Montgomery GW, Ridker PM, Spector TD, Streeten EA, Stefansson K, Thorsteinsdottir U, Uitterlinden AG, Widen E, Murabito JM, Ong KK, Murray A. *Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies*. Nat Genet. 2010 Dec;42(12):1077-85.

Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, Allen HL, Lindgren CM, Luan J, Mägi R, Randall JC, Vedantam S, Winkler TW, Qi L, Workalemahu T, Heid IM, Steinthorsdottir V, Stringham HM, Weedon MN, Wheeler E, Wood AR, Ferreira T, Weyant RJ, Segrè AV, Estrada K, Liang L, Nemesh J, Park JH, Gustafsson S, Kilpeläinen TO, Yang J, Bouatia-Naji N, Esko T, Feitosa MF, Kutalik Z, Mangino M, Raychaudhuri S, Scherag A, Smith AV, Welch R, Zhao JH, Aben KK, Absher DM, **Amin N**, Dixon AL, Fisher E, Glazer NL, Goddard ME, Heard-Costa NL, Hoesel V, Hottenga JJ, Johansson A, Johnson T, Ketkar S, Lamina C, Li S, Moffatt MF, Myers RH, Narisu N, Perry JR, Peters MJ, Preuss M, Ripatti S, Rivadeneira F, Sandholt C, Scott LJ, Timpson NJ, Tyrer JP, van Wingerden S, Watanabe RM, White CC, Wiklund F, Barlassina C, Chasman DI, Cooper MN, Jansson JO, Lawrence RW, Pellikka N, Prokopenko I, Shi J, Thiering E, Alavere H, Alibrandi MT, Almgren P, Arnold AM, Aspelund T, Atwood LD, Balkau B, Balmforth AJ, Bennett AJ, Ben-Shlomo Y, Bergman RN, Bergmann S, Biebermann H, Blakemore AI, Boes T, Bonnycastle LL, Bornstein SR, Brown MJ, Buchanan TA, Busonero F, Campbell H, Cappuccio FP, Cavalcanti-Proença C, Chen YD, Chen CM, Chines PS, Clarke R, Coin L, Connell J, Day IN, Heijer M, Duan J, Ebrahim S, Elliott P, Elosua R, Eiriksdottir G, Erdos MR, Eriksson JG, Facheris MF, Felix SB, Fischer-Posovszky P, Folsom AR, Friedrich N, Freimer NB, Fu M, Gaget S, Gejman PV, Geus EJ, Gieger C, Gjesing AP, Goel A, Goyette P, Grallert H, Grässler J, Greenawalt DM, Groves CJ, Gudnason V, Guiducci C, Hartikainen AL, Hassanali N, Hall AS, Havulinna AS, Hayward C, Heath AC, Hengstenberg C, Hicks AA, Hinney A, Hofman A, Homuth G, Hui J, Igl W, Iribarren C, Isomaa B, Jacobs KB, Jarick I, Jewell E, John U, Jørgensen T, Jousilahti P, Jula A, Kaakinen M, Kajantie E, Kaplan LM, Kathiresan S, Kettunen J, Kinnunen L, Knowles JW, Kolcic I, König IR, Koskinen S, Kovacs P, Kuusisto J, Kraft P, Kvaløy K, Laitinen J, Lantieri O, Lanzani C, Launer LJ, Lecoeur C, Lehtimäki T, Lettre G, Liu J, Lokki ML, Lorentzon M, Luben RN, Ludwig B; MAGIC, Manunta P, Marek D, Marre M, Martin NG, McArdle WL, McCarthy A, McKnight B, Meitinger T, Melander O, Meyre D, Midthjell K, Montgomery GW, Morken MA, Morris AP, Mulic R, Ngwa JS, Nelis M, Neville MJ, Nyholt DR, O'Donnell CJ, O'Rahilly S, Ong KK, Oostra B, Paré G, Parker AN, Perola M, Pichler I, Pietiläinen KH, Platou CG, Polasek O,

Pouta A, Rafelt S, Raitakari O, Rayner NW, Ridderstråle M, Rief W, Ruokonen A, Robertson NR, Rzehak P, Salomaa V, Sanders AR, Sandhu MS, Sanna S, Saramies J, Savolainen MJ, Scherag S, Schipf S, Schreiber S, Schunkert H, Silander K, Sinisalo J, Siscovick DS, Smit JH, Soranzo N, Sovio U, Stephens J, Surakka I, Swift AJ, Tammesoo ML, Tardif JC, Teder-Laving M, Teslovich TM, Thompson JR, Thomson B, Tönjes A, Tuomi T, van Meurs JB, van Ommen GJ, Vatin V, Viikari J, Visvikis-Siest S, Vitart V, Vogel CI, Voight BF, Waite LL, Wallaschofski H, Walters GB, Widen E, Wiegand S, Wild SH, Willemsen G, Witte DR, Witteman JC, Xu J, Zhang Q, Zgaga L, Ziegler A, Zitting P, Beilby JP, Farooqi IS, Hebebrand J, Huikuri HV, James AL, Kähönen M, Levinson DF, Macciardi F, Nieminen MS, Ohlsson C, Palmer LJ, Ridker PM, Stumvoll M, Beckmann JS, Boeing H, Boerwinkle E, Boomsma DI, Caulfield MJ, Chanock SJ, Collins FS, Cupples LA, Smith GD, Erdmann J, Froguel P, Grönberg H, Gyllensten U, Hall P, Hansen T, Harris TB, Hattersley AT, Hayes RB, Heinrich J, Hu FB, Hveem K, Illig T, Jarvelin MR, Kaprio J, Karpe F, Khaw KT, Kiemeney LA, Krude H, Laakso M, Lawlor DA, Metspalu A, Munroe PB, Ouwehand WH, Pedersen O, Penninx BW, Peters A, Pramstaller PP, Quertermous T, Reinehr T, Rissanen A, Rudan I, Samani NJ, Schwarz PE, Shuldiner AR, Spector TD, Tuomilehto J, Uda M, Uitterlinden A, Valle TT, Wabitsch M, Waeber G, Wareham NJ, Watkins H; Procardis Consortium, Wilson JF, Wright AF, Zillikens MC, Chatterjee N, McCarroll SA, Purcell S, Schadt EE, Visscher PM, Assimes TL, Borecki IB, Deloukas P, Fox CS, Groop LC, Haritunians T, Hunter DJ, Kaplan RC, Mohlke KL, O'Connell JR, Peltonen L, Schlessinger D, Strachan DP, van Duijn CM, Wichmann HE, Frayling TM, Thorsteinsdottir U, Abecasis GR, Barroso I, Boehnke M, Stefansson K, North KE, McCarthy MI, Hirschhorn JN, Ingelsson E, Loos RJ. *Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index*. Nat Genet. 2010 Nov;42(11):937-48.

Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V, Thorleifsson G, Zillikens MC, Speliotes EK, Mägi R, Workalemahu T, White CC, Bouatia-Naji N, Harris TB, Berndt SI, Ingelsson E, Willer CJ, Weedon MN, Luan J, Vedantam S, Esko T, Kilpeläinen TO, Kutalik Z, Li S, Monda KL, Dixon AL, Holmes CC, Kaplan LM, Liang L, Min JL, Moffatt MF, Molony C, Nicholson G, Schadt EE, Zondervan KT, Feitosa MF, Ferreira T, Allen HL, Weyant RJ, Wheeler E, Wood AR; MAGIC, Estrada K, Goddard ME, Lettre G, Mangino M, Nyholt DR, Purcell S, Smith AV, Visscher PM, Yang J, McCarroll SA, Nemesh J, Voight BF, Absher D, **Amin N**, Aspelund T, Coin L, Glazer NL, Hayward C, Heard-Costa NL, Hottenga JJ, Johansson A, Johnson T, Kaakinen M, Kapur K, Ketkar S, Knowles JW, Kraft P, Kraja AT, Lamina C, Leitzmann MF, McKnight B, Morris AP, Ong KK, Perry JR, Peters MJ, Polasek O, Prokopenko I, Rayner NW, Ripatti S, Rivadeneira F, Robertson NR, Sanna S, Sovio U, Surakka I, Teumer A, van Wingerden S, Vitart V, Zhao JH, Cavalcanti-Proença C, Chines PS, Fisher E, Kulzer JR, Lecoeur C, Narisu N, Sandholt C, Scott LJ, Silander K, Stark K, Tammesoo ML, Teslovich TM, Timpson NJ, Watanabe RM, Welch R, Chasman DI, Cooper MN, Jansson JO, Kettunen J, Lawrence RW, Pellikka N, Perola M, Vandenput L, Alavere H, Almgren P, Atwood LD, Bennett AJ, Biffar R, Bonnycastle LL, Bornstein SR, Buchanan TA, Campbell H, Day IN, Dei M, Dörr M, Elliott P, Erdos MR, Eriksson JG, Freimer NB, Fu M, Gaget S, Geus EJ, Gjesing AP, Grallert H, Grässler J, Groves CJ,

Guiducci C, Hartikainen AL, Hassanali N, Havulinna AS, Herzig KH, Hicks AA, Hui J, Igl W, Jousilahti P, Jula A, Kajantie E, Kinnunen L, Kolcic I, Koskinen S, Kovacs P, Kroemer HK, Krzelj V, Kuusisto J, Kvaloy K, Laitinen J, Lantieri O, Lathrop GM, Lokki ML, Luben RN, Ludwig B, McArdle WL, McCarthy A, Morken MA, Nelis M, Neville MJ, Paré G, Parker AN, Peden JF, Pichler I, Pietiläinen KH, Platou CG, Pouta A, Ridderstråle M, Samani NJ, Saramies J, Sinisalo J, Smit JH, Strawbridge RJ, Stringham HM, Swift AJ, Teder-Laving M, Thomson B, Usala G, van Meurs JB, van Ommen GJ, Vatin V, Volpato CB, Wallaschofski H, Walters GB, Widen E, Wild SH, Willemsen G, Witte DR, Zgaga L, Zitting P, Beilby JP, James AL, Kähönen M, Lehtimäki T, Nieminen MS, Ohlsson C, Palmer LJ, Raitakari O, Ridker PM, Stumvoll M, Tönjes A, Viikari J, Balkau B, Ben-Shlomo Y, Bergman RN, Boeing H, Smith GD, Ebrahim S, Froguel P, Hansen T, Hengstenberg C, Hveem K, Isomaa B, Jørgensen T, Karpe F, Khaw KT, Laakso M, Lawlor DA, Marre M, Meitinger T, Metspalu A, Midthjell K, Pedersen O, Salomaa V, Schwarz PE, Tuomi T, Tuomilehto J, Valle TT, Wareham NJ, Arnold AM, Beckmann JS, Bergmann S, Boerwinkle E, Boomsma DI, Caulfield MJ, Collins FS, Eiriksdottir G, Gudnason V, Gyllensten U, Hamsten A, Hattersley AT, Hofman A, Hu FB, Illig T, Iribarren C, Jarvelin MR, Kao WH, Kaprio J, Launer LJ, Munroe PB, Oostra B, Penninx BW, Pramstaller PP, Psaty BM, Quertermous T, Rissanen A, Rudan I, Shuldiner AR, Soranzo N, Spector TD, Syvanen AC, Uda M, Uitterlinden A, Völzke H, Vollenweider P, Wilson JF, Witteman JC, Wright AF, Abecasis GR, Boehnke M, Borecki IB, Deloukas P, Frayling TM, Groop LC, Haritunians T, Hunter DJ, Kaplan RC, North KE, O'Connell JR, Peltonen L, Schlessinger D, Strachan DP, Hirschhorn JN, Assimes TL, Wichmann HE, Thorsteinsdottir U, van Duijn CM, Stefansson K, Cupples LA, Loos RJ, Barroso I, McCarthy MI, Fox CS, Mohlke KL, Lindgren CM. *Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution*. Nat Genet. 2010 Nov;42(11):949-60.

Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, Ferreira T, Wood AR, Weyant RJ, Segrè AV, Speliotes EK, Wheeler E, Soranzo N, Park JH, Yang J, Gudbjartsson D, Heard-Costa NL, Randall JC, Qi L, Vernon Smith A, Mägi R, Pastinen T, Liang L, Heid IM, Luan J, Thorleifsson G, Winkler TW, Goddard ME, Sin Lo K, Palmer C, Workalemahu T, Aulchenko YS, Johansson A, Zillikens MC, Feitosa MF, Esko T, Johnson T, Ketkar S, Kraft P, Mangino M, Prokopenko I, Absher D, Albrecht E, Ernst F, Glazer NL, Hayward C, Hottenga JJ, Jacobs KB, Knowles JW, Kutalik Z, Monda KL, Polasek O, Preuss M, Rayner NW, Robertson NR, Steinthorsdottir V, Tyrer JP, Voight BF, Wiklund F, Xu J, Hua Zhao J, Nyholt DR, Pellikka N, Perola M, Perry JR, Surakka I, Tammesoo ML, Altmaier EL, **Amin N**, Aspelund T, Bhangale T, Boucher G, Chasman DI, Chen C, Coin L, Cooper MN, Dixon AL, Gibson Q, Grundberg E, Hao K, Juhani Junttila M, Kaplan LM, Kettunen J, König IR, Kwan T, Lawrence RW, Levinson DF, Lorentzon M, McKnight B, Morris AP, Müller M, Suh Ngwa J, Purcell S, Rafelt S, Salem RM, Salvi E, Sanna S, Shi J, Sovio U, Thompson JR, Turchin MC, Vandenput L, Verlaan DJ, Vitart V, White CC, Ziegler A, Almgren P, Balmforth AJ, Campbell H, Citterio L, De Grandi A, Dominiczak A, Duan J, Elliott P, Elosua R, Eriksson JG, Freimer NB, Geus EJ, Glorioso N, Haiqing S, Hartikainen AL,

Havulinna AS, Hicks AA, Hui J, Igl W, Illig T, Jula A, Kajantie E, Kilpeläinen TO, Koiranen M, Kolcic I, Koskinen S, Kovacs P, Laitinen J, Liu J, Lokki ML, Marusic A, Maschio A, Meitinger T, Mulas A, Paré G, Parker AN, Peden JF, Petersmann A, Pichler I, Pietiläinen KH, Pouta A, Ridderstråle M, Rotter JI, Sambrook JG, Sanders AR, Oliver Schmidt C, Sinisalo J, Smit JH, Stringham HM, Bragi Walters G, Widen E, Wild SH, Willemsen G, Zagato L, Zgaga L, Zitting P, Alavere H, Farrall M, McArdle WL, Nelis M, Peters MJ, Ripatti S, van Meurs JB, Aben KK, Ardlie KG, Beckmann JS, Beilby JP, Bergman RN, Bergmann S, Collins FS, Cusi D, den Heijer M, Eiriksdottir G, Gejman PV, Hall AS, Hamsten A, Huikuri HV, Iribarren C, Kähönen M, Kaprio J, Kathiresan S, Kiemeney L, Kocher T, Launer LJ, Lehtimäki T, Melander O, Mosley TH Jr, Musk AW, Nieminen MS, O'Donnell CJ, Ohlsson C, Oostra B, Palmer LJ, Raitakari O, Ridker PM, Rioux JD, Rissanen A, Rivolta C, Schunkert H, Shuldiner AR, Siscovick DS, Stumvoll M, Tönjes A, Tuomilehto J, van Ommen GJ, Viikari J, Heath AC, Martin NG, Montgomery GW, Province MA, Kayser M, Arnold AM, Atwood LD, Boerwinkle E, Chanock SJ, Deloukas P, Gieger C, Grönberg H, Hall P, Hattersley AT, Hengstenberg C, Hoffman W, Mark Lathrop G, Salomaa V, Schreiber S, Uda M, Waterworth D, Wright AF, Assimes TL, Barroso I, Hofman A, Mohlke KL, Boomsma DI, Caulfield MJ, Adrienne Cupples L, Erdmann J, Fox CS, Gudnason V, Gyllensten U, Harris TB, Hayes RB, Jarvelin MR, Mooser V, Munroe PB, Ouwehand WH, Penninx BW, Pramstaller PP, Quertermous T, Rudan I, Samani NJ, Spector TD, Völzke H, Watkins H, Wilson JF, Groop LC, Haritunians T, Hu FB, Kaplan RC, Metspalu A, North KE, Schlessinger D, Wareham NJ, Hunter DJ, O'Connell JR, Strachan DP, Wichmann HE, Borecki IB, van Duijn CM, Schadt EE, Thorsteinsdottir U, Peltonen L, Uitterlinden AG, Visscher PM, Chatterjee N, Loos RJ, Boehnke M, McCarthy MI, Ingelsson E, Lindgren CM, Abecasis GR, Stefansson K, Frayling TM, Hirschhorn JN. *Hundreds of variants clustered in genomic loci and biological pathways affect human height*. Nature. 2010 Oct 14;467(7317):832-8.

Solouki AM, Verhoeven VJ, van Duijn CM, Verkerk AJ, Ikram MK, Hysi PG, Despriet DD, van Koolwijk LM, Ho L, Ramdas WD, Czudowska M, Kuijpers RW, **Amin N**, Struchalin M, Aulchenko YS, van Rij G, Riemslag FC, Young TL, Mackey DA, Spector TD, Gorgels TG, Willemse-Assink JJ, Isaacs A, Kramer R, Swagemakers SM, Bergen AA, van Oosterhout AA, Oostra BA, Rivadeneira F, Uitterlinden AG, Hofman A, de Jong PT, Hammond CJ, Vingerling JR, Klaver CC. *A genome-wide association study identifies a susceptibility locus for refractive errors and myopia at 15q14*. Nat Genet. 2010 Oct;42(10):897-901.

Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G, McCulloch LJ, Ferreira T, Grallert H, **Amin N**, Wu G, Willer CJ, Raychaudhuri S, McCarroll SA, Langenberg C, Hofmann OM, Dupuis J, Qi L, Segrè AV, van Hoek M, Navarro P, Ardlie K, Balkau B, Benediktsson R, Bennett AJ, Blagieva R, Boerwinkle E, Bonnycastle LL, Bengtsson Boström K, Bravenboer B, Bumpstead S, Burtt NP, Charpentier G, Chines PS, Cornelis M, Couper DJ, Crawford G, Doney AS, Elliott KS, Elliott AL, Erdos MR, Fox CS, Franklin CS, Ganser

M, Gieger C, Grarup N, Green T, Griffin S, Groves CJ, Guiducci C, Hadjadj S, Hassanali N, Herder C, Isomaa B, Jackson AU, Johnson PR, Jørgensen T, Kao WH, Klopp N, Kong A, Kraft P, Kuusisto J, Lauritzen T, Li M, Lieverse A, Lindgren CM, Lyssenko V, Marre M, Meitinger T, Midthjell K, Morken MA, Narisu N, Nilsson P, Owen KR, Payne F, Perry JR, Petersen AK, Platou C, Proença C, Prokopenko I, Rathmann W, Rayner NW, Robertson NR, Rocheleau G, Roden M, Sampson MJ, Saxena R, Shields BM, Shrader P, Sigurdsson G, Sparsø T, Strassburger K, Stringham HM, Sun Q, Swift AJ, Thorand B, Tichet J, Tuomi T, van Dam RM, van Haeften TW, van Herpt T, van Vliet-Ostaptchouk JV, Walters GB, Weedon MN, Wijmenga C, Witteman J, Bergman RN, Cauchi S, Collins FS, Gloyn AL, Gyllensten U, Hansen T, Hide WA, Hitman GA, Hofman A, Hunter DJ, Hveem K, Laakso M, Mohlke KL, Morris AD, Palmer CN, Pramstaller PP, Rudan I, Sijbrands E, Stein LD, Tuomilehto J, Uitterlinden A, Walker M, Wareham NJ, Watanabe RM, Abecasis GR, Boehm BO, Campbell H, Daly MJ, Hattersley AT, Hu FB, Meigs JB, Pankow JS, Pedersen O, Wichmann HE, Barroso I, Florez JC, Frayling TM, Groop L, Sladek R, Thorsteinsdottir U, Wilson JF, Illig T, Froguel P, van Duijn CM, Stefansson K, Altshuler D, Boehnke M, McCarthy MI; MAGIC investigators; GIANT Consortium. *Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis*. Nat Genet. 2010 Jul;42(7):579-89.

Demirkan A, Penninx BW, Hek K, Wray NR, **Amin N**, Aulchenko YS, van Dyck R, de Geus EJ, Hofman A, Uitterlinden AG, Hottenga JJ, Nolen WA, Oostra BA, Sullivan PF, Willemsen G, Zitman FG, Tiemeier H, Janssens AC, Boomsma DI, van Duijn CM, Middeldorp CM. *Genetic risk profiles for depression and anxiety in adult and elderly cohorts*. Mol Psychiatry. 2010 Jun 22.

Ramdas WD, van Koolwijk LM, Ikram MK, Jansonius NM, de Jong PT, Bergen AA, Isaacs A, **Amin N**, Aulchenko YS, Wolfs RC, Hofman A, Rivadeneira F, Oostra BA, Uitterlinden AG, Hysi P, Hammond CJ, Lemij HG, Vingerling JR, Klaver CC, van Duijn CM. A *genome-wide association study of optic disc parameters.* PLoS Genet. 2010 Jun 10;6(6)

Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, **Amin N**, Geller F, Sulem P, Rafnar T, Esko T, Walter S, Gieger C, Rawal R, Mangino M, Prokopenko I, Mägi R, Keskitalo K, Gudjonsdottir IH, Gretarsdottir S, Stefansson H, Thompson JR, Aulchenko YS, Nelis M, Aben KK, den Heijer M, Dirksen A, Ashraf H, Soranzo N, Valdes AM, Steves C, Uitterlinden AG, Hofman A, Tönjes A, Kovacs P, Hottenga JJ, Willemsen G, Vogelzangs N, Döring A, Dahmen N, Nitz B, Pergadia ML, Saez B, De Diego V, Lezcano V, Garcia-Prats MD, Ripatti S, Perola M, Kettunen J, Hartikainen AL, Pouta A, Laitinen J, Isohanni M, Huei-Yi S, Allen M, Krestyaninova M, Hall AS, Jones GT, van Rij AM, Mueller T, Dieplinger B, Haltmayer M, Jonsson S, Matthiasson SE, Oskarsson H, Tyrfingsson T, Kiemeney LA, Mayordomo JI, Lindholt JS, Pedersen JH, Franklin WA, Wolf H, Montgomery GW, Heath AC, Martin NG, Madden PA, Giegling I, Rujescu D, Järvelin MR, Salomaa V, Stumvoll M, Spector TD, Wichmann HE, Metspalu A, Samani NJ, Penninx BW, Oostra BA, Boomsma DI, Tiemeier H, van Duijn CM, Kaprio J, Gulcher JR; ENGAGE Consortium, McCarthy MI, Peltonen L, Thorsteinsdottir

U, Stefansson K. *Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior*. Nat Genet. 2010 May;42(5):448-53.

Heid IM, Henneman P, Hicks A, Coassin S, Winkler T, Aulchenko YS, Fuchsberger C, Song K, Hivert MF, Waterworth DM, Timpson NJ, Richards JB, Perry JR, Tanaka T, **Amin N**, Kollerits B, Pichler I, Oostra BA, Thorand B, Frants RR, Illig T, Dupuis J, Glaser B, Spector T, Guralnik J, Egan JM, Florez JC, Evans DM, Soranzo N, Bandinelli S, Carlson OD, Frayling TM, Burling K, Smith GD, Mooser V, Ferrucci L, Meigs JB, Vollenweider P, Dijk KW, Pramstaller P, Kronenberg F, van Duijn CM. *Clear detection of ADIPOQ locus as the major gene for plasma adiponectin: results of genome-wide association analyses including 4659 European individuals*. Atherosclerosis. 2010 Feb;208(2):412-20.

Kong A, Steinthorsdottir V, Masson G, Thorleifsson G, Sulem P, Besenbacher S, Jonasdottir A, Sigurdsson A, Kristinsson KT, Jonasdottir A, Frigge ML, Gylfason A, Olason PI, Gudjonsson SA, Sverrisson S, Stacey SN, Sigurgeirsson B, Benediktsdottir KR, Sigurdsson H, Jonsson T, Benediktsson R, Olafsson JH, Johannsson OT, Hreidarsson AB, Sigurdsson G; **DIAGRAM Consortium**, Ferguson-Smith AC, Gudbjartsson DF, Thorsteinsdottir U, Stefansson K. *Parental origin of sequence variants associated with complex diseases*. Nature. 2009 Dec 17;462(7275):868-74.

Kettunen J, Silander K, Saarela O, **Amin N**, Müller M, Timpson N, Surakka I, Ripatti S, Laitinen J, Hartikainen AL, Pouta A, Lahermo P, Anttila V, Männistö S, Jula A, Virtamo J, Salomaa V, Lehtimäki T, Raitakari O, Gieger C, Wichmann EH, Van Duijn CM, Smith GD, McCarthy MI, Järvelin MR, Perola M, Peltonen L. *European lactase persistence genotype shows evidence of association with increase in body mass index*. Hum Mol Genet. 2010 Mar 15;19(6):1129-36.

Richards JB, Kavvoura FK, Rivadeneira F, Styrkársdóttir U, Estrada K, Halldórsson BV, Hsu YH, Zillikens MC, Wilson SG, Mullin BH, **Amin N**, Aulchenko YS, Cupples LA, Deloukas P, Demissie S, Hofman A, Kong A, Karasik D, van Meurs JB, Oostra BA, Pols HA, Sigurdsson G, Thorsteinsdottir U, Soranzo N, Williams FM, Zhou Y, Ralston SH, Thorleifsson G, van Duijn CM, Kiel DP, Stefansson K, Uitterlinden AG, Ioannidis JP, Spector TD; Genetic Factors for Osteoporosis Consortium. *Collaborative meta-analysis: associations of 150 candidate genes with osteoporosis and osteoporotic fracture*. Ann Intern Med. 2009 Oct 20;151(8):528-37.

Rivadeneira F, Styrkársdottir U, Estrada K, Halldórsson BV, Hsu YH, Richards JB, Zillikens MC, Kavvoura FK, **Amin N**, Aulchenko YS, Cupples LA, Deloukas P, Demissie S, Grundberg E, Hofman A, Kong A, Karasik D, van Meurs JB, Oostra B, Pastinen T, Pols HA, Sigurdsson G, Soranzo N, Thorleifsson G, Thorsteinsdottir U, Williams FM, Wilson SG, Zhou Y, Ralston SH, van Duijn CM, Spector T, Kiel DP, Stefansson K, Ioannidis JP, Uitterlinden AG; Genetic Factors for Osteoporosis (GEFOS) Consortium. *Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies*. Nat Genet. 2009 Nov;41(11):1199-206. Epub 2009 Oct 4.

Zillikens MC, van Meurs JB, Rivadeneira F, **Amin N**, Hofman A, Oostra BA, Sijbrands EJ, Witteman JC, Pols HA, van Duijn CM, Uitterlinden AG. *SIRT1 genetic variation is related to BMI and risk of obesity*. Diabetes. 2009 Dec;58(12):2828-34. Epub 2009 Sep 9.

Heard-Costa NL, Zillikens MC, Monda KL, Johansson A, Harris TB, Fu M, Haritunians T, Feitosa MF, Aspelund T, Eiriksdottir G, Garcia M, Launer LJ, Smith AV, Mitchell BD, McArdle PF, Shuldiner AR, Bielinski SJ, Boerwinkle E, Brancati F, Demerath EW, Pankow JS, Arnold AM, Chen YD, Glazer NL, McKnight B, Psaty BM, Rotter JI, **Amin N**, Campbell H, Gyllensten U, Pattaro C, Pramstaller PP, Rudan I, Struchalin M, Vitart V, Gao X, Kraja A, Province MA, Zhang Q, Atwood LD, Dupuis J, Hirschhorn JN, Jaquish CE, O'Donnell CJ, Vasan RS, White CC, Aulchenko YS, Estrada K, Hofman A, Rivadeneira F, Uitterlinden AG, Witteman JC, Oostra BA, Kaplan RC, Gudnason V, O'Connell JR, Borecki IB, van Duijn CM, Cupples LA, Fox CS, North KE. *NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium*. PLoS Genet. 2009 Jun;5(6).

Lindgren CM, Heid IM, Randall JC, Lamina C, Steinthorsdottir V, Qi L, Speliotes EK, Thorleifsson G, Willer CJ, Herrera BM, Jackson AU, Lim N, Scheet P, Soranzo N, **Amin N**, Aulchenko YS, Chambers JC, Drong A, Luan J, Lyon HN, Rivadeneira F, Sanna S, Timpson NJ, Zillikens MC, Zhao JH, Almgren P, Bandinelli S, Bennett AJ, Bergman RN, Bonnycastle LL, Bumpstead SJ, Chanock SJ, Cherkas L, Chines P, Coin L, Cooper C, Crawford G, Doering A, Dominiczak A, Doney AS, Ebrahim S, Elliott P, Erdos MR, Estrada K, Ferrucci L, Fischer G, Forouhi NG, Gieger C, Grallert H, Groves CJ, Grundy S, Guiducci C, Hadley D, Hamsten A, Havulinna AS, Hofman A, Holle R, Holloway JW, Illig T, Isomaa B, Jacobs LC, Jameson K, Jousilahti P, Karpe F, Kuusisto J, Laitinen J, Lathrop GM, Lawlor DA, Mangino M, McArdle WL, Meitinger T, Morken MA, Morris AP, Munroe P, Narisu N, Nordström A, Nordström P, Oostra BA, Palmer CN, Payne F, Peden JF, Prokopenko I, Renström F, Ruokonen A, Salomaa V, Sandhu MS, Scott LJ, Scuteri A, Silander K, Song K, Yuan X, Stringham HM, Swift AJ, Tuomi T, Uda M, Vollenweider P, Waeber G, Wallace C, Walters GB, Weedon MN; Wellcome Trust Case Control Consortium, Witteman JC, Zhang C, Zhang W, Caulfield MJ, Collins FS, Davey Smith G, Day IN, Franks PW, Hattersley AT, Hu FB, Jarvelin MR, Kong A, Kooner JS, Laakso M, Lakatta E, Mooser V, Morris AD, Peltonen L, Samani NJ, Spector TD, Strachan DP, Tanaka T, Tuomilehto J, Uitterlinden AG, van Duijn CM, Wareham NJ, Hugh Watkins; Procardis Consortia, Waterworth DM, Boehnke M, Deloukas P, Groop L, Hunter DJ, Thorsteinsdottir U, Schlessinger D, Wichmann HE, Frayling TM, Abecasis GR, Hirschhorn JN, Loos RJ, Stefansson K, Mohlke KL, Barroso I, McCarthy MI; Giant Consortium. *Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution*. PLoS Genet. 2009 Jun;5(6)

Vink JM, Smit AB, de Geus EJ, Sullivan P, Willemsen G, Hottenga JJ, Smit JH, Hoogendijk WJ, Zitman FG, Peltonen L, Kaprio J, Pedersen NL, Magnusson PK, Spector TD, Kyvik KO, Morley KI,

Heath AC, Martin NG, Westendorp RG, Slagboom PE, Tiemeier H, Hofman A, Uitterlinden AG, Aulchenko YS, **Amin N**, van Duijn C, Penninx BW, Boomsma DI. *Genome-wide association study of smoking initiation and current smoking*. Am J Hum Genet. 2009 Mar;84(3):367-79.

Terracciano A, Sanna S, Uda M, Deiana B, Usala G, Busonero F, Maschio A, Scally M, Patriciu N, Chen WM, Distel MA, Slagboom EP, Boomsma DI, Villafuerte S, Sliwerska E, Burmeister M, **Amin N**, Janssens AC, van Duijn CM, Schlessinger D, Abecasis GR, Costa PT Jr. *Genome-wide association scan for five major dimensions of personality*. Mol Psychiatry. 2010 Jun;15(6):647-56.

Najaf Amin, Maaike Schuur, Elena S. Gusareva, Aaron J. Isaacs, Yurii S. Aulchenko, Anatoly V. Kirichenko, Irina V. Zorkoltseva, Tatiana I. Axenovich, Ben A. Oostra, A. Cecile J.W. Janssens, Cornelia M. van Duijn. *A genome wide linkage study of individuals with high scores on NEO personality traits*. (Submitted)

Karla V. Allebrandt, **Najaf Amin**, Bertram Müller-Myhsok, Tõnu Esko, Maris Teder-Laving, Renata V.D.M. Azevedo, Caroline Hayward, Josine van Mill, Nicole Vogelzangs, Edward W. Green, Scott A. Melville, H.-Erich Wichmann, Ben A. Oostra, A. Cecile J.W. Janssens, Harry Campbell, James F. Wilson, Andrew A. Hicks, Peter P. Pramstaller, Zoran Dogas, Igor Rudan, Martha Merrow, Brenda Penninx, Charalambos P. Kyriacou, Andres Metspalu, Cornelia M. van Duijn, Thomas Meitinger and Till Roenneberg. *ABCC9* modulates sleep duration in humans and in *Drosophila*. (Submitted)

Najaf Amin, Viatcheslav Saviouk, Narelle K. Hansell, A. Cecile J.W. Janssens, Marleen H.M. de Moor, Pamela A. F. Madden, Irina V. Zorkoltseva, Brenda W. Penninx, Antonio Terracciano, Manuela Uda, Toshiko Tanaka, Tonu Esko, Anu Realo, Luigi Ferrucci, Michelle Luciano, Gail Davies, Andres Metspalu, Goncalo R. Abecasis, Ian J. Deary, Katri Raikkonen, Laura J. Bierut, Paul T. Costa, Jouke J. Hottenga, Gu Zhu, Anatoly V. Kirichenko, Aaron J. Isaacs, Yurii S. Aulchenko, Gonneke Willemsen, Andrew C. Heath, and Michele L Pergadia, Sarah E. Medland, Tatiana I. Axenovich, Eco de Geus, Grant W. Montgomery, Margaret J. Wright, Ben A. Oostra, Nicholas G. Martin, Dorret I. Boomsma, Cornelia M. van Duijn. *Refining genome-wide linkage intervals using GWAS identifies several loci influencing NEO Personality Dimensions.* (Submitted)

Najaf Amin, Karla V. Allebrandt, Bertram Müller-Myhsok, Karin Hek, Maris Teder-Laving, Caroline Hayward, Tõnu Esko, Josine G. van Mill, Scott A. Melville, Nicole Vogelzangs, Harry Campbell, Ayse Demirkan, Zoran Dogas, Fran Borovecki, Jennifer E.Huffman, Yurii S. Aulchenko, Andre G. Uitterlinden, Eric Wichmann, Albert Hofman, Ben A. Oostra, Alan F. Wright, James F. Wilson, Peter P. Pramstaller, Andrew A. Hicks, A. Cecile J.W. Janssens, Martha Merrow, Igor Rudan, Henning Tiemeier, Brenda W. Penninx, Andres Metspalu, Thomas Meitinger, Till Roenneberg, and Cornelia M. van Duijn. *Common genetic variant in HRNBP3 is associated with sleep latency*. (Submitted).

Najaf Amin, Enda Byrne, Julie Johnson, Georgia Chenevix-Trench, Stefan Walter, Ilja M. Nolte, kConFab Investigators, Jacqueline M. Vink, Rajesh Rawal, Massimo Mangino, Alexander Teumer, Joost C. Keers, Germaine Verwoert, Sebastian Baumeister, Reiner Biffar, Astrid Petersmann, Norbert Dahmen, Angela Doering, Aaron J. Isaacs, Linda Broer, Naomi R. Wray, Grant W. Montgomery, Daniel Levy, Bruce M. Psaty, Vilmundur Gudnason, Aravinda Chakravarti, Unnur Thorsteinsdottir, Frank J.A. van Rooij, Yurii S. Aulchenko, Jouke J. Hottenga, Fernando R. Rivadeneira, Albert Hofman, Andre G. Uitterlinden, Chris J. Hammond, So-Youn Shin, Arfan Ikram, Jacqueline C.M. Witteman, A. Cecile J.W. Janssens, Harold Snieder, Henning Tiemier, Bruce H. R. Wolfenbuttel, Ben A. Oostra, Andrew C. Heath, Erich Wichmann, Tim D. Spector, Hans J. Grabe, Dorret I. Boomsma, Nicholas G. Martin, Cornelia M. van Duijn. *Genome-wide association and gene expression analyses suggest association of CYP1A1/CYP1A2 with coffee drinking.* (Submitted).