Keeping focus:
a study on Attention Problems in the GWAS era

M.M. Groen-Blokhuis

Table of contents

Chapter 1

General Introduction

This thesis addresses questions on the development of children, with an important focus on Attention Problems (AP) and Attention Deficit Hyperactivity Disorder (ADHD) in preschool and school-age children. The research described in this thesis was conducted between September 2009 and December 2013, a period in which genome-wide association studies (GWAS) became successful and were dominating the high-impact journals. As large sample sizes are a prerequisite for success in this type of study and collaboration very rapidly became the norm in genetics, a considerable part of my PhD trajectory has been dedicated to GWA studies for large consortia, such as the EArly Genetics and Lifecourse Epidemiology (EAGLE) and Early Growth Genetics (EGG) consortium. Alongside this consortium work, I spent time on the analysis of specific genetic and environmental risk factors for behavioral problems and AP in particular, using data from the Netherlands Twin Register (NTR). In the last year of my PhD trajectory, these two lines of research came together in the opportunity to lead a consortium GWAS meta-analysis on continuous measures of ADHD. Although the scope of this thesis is somewhat broader; this introduction will focus on the etiology of AP and ADHD in childhood.

AP and ADHD are characterized by age-inappropriate inattentive, hyperactive and impulsive behaviors that persist across settings, i.e. are present both at home and in the school environment.[1] The prevalence of a clinical diagnosis of ADHD is estimated at 5.9-7.2% in childhood and adolescence, and is similar across countries and regions of the world.[2] Data on the socio-economic impact of ADHD, although somewhat sparse, suggest that ADHD is an economic and social burden to society.[3, 4] Treatment options such as stimulant medication and, to a lesser extent, behavioral therapy show benefits in terms of symptom reduction in ADHD, yet, no cure is available.[5, 6] Children diagnosed with ADHD show a reduced quality of life and substantial impairment in both social and academic functioning[7, 8] and childhood ADHD is associated with various adverse long-term outcomes such as lower educational attainment, lower occupational status, substance abuse, risk behavior and criminality.[3, 9, 10] However, these data should not be interpreted as if the impairment in children diagnosed with ADHD is a fixed reality; to some extent the impairment is also context dependent. Characteristics that are a burden in e.g. the school system can be advantageous in particular settings; impulsivity, hyperactivity and a rapid shift of attention were probably adaptive in the past, when high risk environments were more common.[11, 12] Nowadays, these traits can still be beneficial in risk-taking

occupations, as indicated by the fact that entrepreneurial careers are more common among adults that score high on ADHD symptom levels.[13, 14]

A clear causal agent for ADHD and ADHD-like behaviors is lacking and the pathophysiology underlying the disorder is unknown. ADHD, AP and other behavioral disorders are thought to be complex multifactorial traits, with many genetic and environmental factors contributing to their development. Identifying specific genetic and environmental risk factors for AP and ADHD is important for clarification of the underlying etiology, refinement of the diagnostic system, and the development of better therapies for children suffering from ADHD. Here, a summary of the literature on this topic is provided, starting with a discussion on measurement issues of the trait.

Measurement of AP and ADHD. In ADHD research, two different types of outcome measures can be distinguished. Some studies employ a categorical framework using diagnostic criteria from manuals used in clinical practice, such as ICD-10, DSM-IV, and its recent update DSM-5.[1, 15, 16] However, ADHD can also be viewed as the extreme end of a continuous distribution of inattentive and hyperactive behaviors that can be observed in the general population.[17] Several measurement instruments have been developed to assess ADHD-like and other problem behaviors on a continuous scale. Whereas some of these scales have based their item content on the DSM-IV manual, like for example the Strengths and Weaknesses of ADHD Symptoms and Normal Behavior Scale (SWAN),[18] other have used a bottom-up approach and performed principal components analyses on item level data to derive syndrome scales, e.g. the Child Behavior Checklist (CBCL) and the Strengths and Difficulties Questionnaire (SDQ).[19-21] Although these scales differ in their exact item content, all of them contain a syndrome scale that is to a considerable extent related to a clinical diagnosis of ADHD.[22, 23] Yet, the overlap is less than perfect and the question remains whether ADHD is best understood as a continuous or as a categorical trait. Studies that assessed this question have often employed statistical methods such as Latent Class Analysis and Factor Mixture Modeling on item level data, but these studies showed somewhat mixed results.[24-28] Another approach to address this question is to assess whether risk factors for ADHD have an effect across the full range of behavior in the population, as this would strongly support the use of dimensional models. This could have important implications; although most clinicians are aware of the possible continuous nature of ADHD, dimensional models are only rarely implemented systematically in clinical practice. In addition,

many research projects are focused on the analyses of cases versus controls, and do not include the full information available from symptom counts.

Genetic risk factors. As with most other psychiatric disorders, there is an important contribution of genes to individual differences in AP and ADHD, explaining why these traits run in families, but also why siblings in the same family may differ from each other. Irrespective whether a clinical diagnosis of ADHD or a continuous measure of ADHD behaviors is studied, twin, adoption, and family studies have consistently found childhood ADHD to be highly heritable in childhood, with a mean heritability estimate of 76% across twenty studies in western countries.[29-35]

Given the high heritability of ADHD, many studies have tried to identify specific genetic variants underlying the genetic risk for ADHD. Initial molecular genetic studies have focused on genes that were selected based on existent neurobiological knowledge. Some of these studies were successful; a meta-analysis of candidate gene studies in ADHD found evidence for the implication of genes in dopaminergic, serotonergic and neurosystem development pathways in ADHD, specifically for variants in DAT1, DRD4, DRD5, 5HTT, HTR1B, and SNAP25.[36] However, the above listed variants still explain only a small amount of variance and ADHD is somewhat exceptional in finding replicable associations using a candidate gene approach. As a whole, this approach has not been very fruitful for psychiatric disorders: most initial findings have not been replicated and were probably either spurious or of inflated effect size. With the costs of large-scale genotyping dropping over time, hypothesis-free methods have become feasible in the last decade. The first successful genome-wide association study (GWAS) of age-related macular degeneration was published in 2005;[37] and the first psychiatric GWAS was published in 2007; since then, the number of published GWAS has grown rapidly.[38] For a long time, the successes of GWAS on psychiatric phenotypes have been lacking somewhat behind, and no genome-wide significant findings were reported in the largest GWA meta-analysis on ADHD so far.[39-41] The latter may be explained by the relatively small sample size (5,621 ADHD cases and 13,589 controls since the last update), however, almost all GWAS projects on non-psychiatric phenotypes have found genome-wide significant results when including > 11,000 subjects, suggesting that these null findings might be due to a larger degree of phenotypic and genotypic heterogeneity in ADHD and psychiatric disorders in general.[41] Overall, the variants

detected in GWAS on both psychiatric and non-psychiatric phenotypes explain only a small proportion of variation. This gap between high heritability estimates and a low amount of variance explained by detected genetic variants has been referred to as the missing heritability problem.[42] It has been suggested that studies in which family members (e.g. twins) were rated by the same observer (same parent, same teacher) estimate heritability to be larger than studies in which heritability is estimated from individually rated participants, as is usually the case in genetic studies of unrelated subjects who take part in GWA studies.[43] Several other explanations for the missing heritability have been proposed, including an important contribution of rare variants, many common variants of very small effect size, and gene-gene or gene-environment interactions.[42] GWA studies typically investigate only common variation of a single base pair, called single nucleotide polymorphisms (SNPs), where common is generally referring to variants with a frequency of 5 or 1% in the population. However, there is much more variation present in the genome, including small insertions and deletions, referred to as indels, and larger insertions and deletions named copy number variants (CNVs). Currently, empirical evidence supports at least a contribution of rare variants and common SNPs with small effect size to ADHD. It has been known for some time that children with rare chromosomal microdeletion syndromes have an increased risk to develop ADHD and other psychiatric disorders.[44] In addition, a higher burden of large (> 500 kb) rare (< 1%) CNVs has been found for ADHD cases compared to controls.[45, 46] Evidence for the contribution of common SNPs has come from methods that investigate the variation explained by all SNPs included on the arrays. The Genetic Complex Trait Analysis (GCTA) software package uses a linear mixed model approach in which the observed phenotypic similarities of all possible pairs of individuals are regressed on their observed genetic similarities.[47] So et al. have introduced a different method that addresses the same question by comparing the observed distribution of effect sizes to the expected distribution.[48] Studies employing these methods suggest that for many phenotypes such as height, weight, smoking or IQ, roughly one-third to half of the genetic variation estimated in twin studies can be attributed to the effects of SNP on current GWAS platforms.[47, 49-51] Psychiatric disorders are no exception here; the Psychiatric Genomics Consortium (PGC) reported GCTA estimates for major schizophrenia, bipolar disorder, major depressive disorder, autism spectrum disorders and ADHD that range between 21 and 57% of the genetic variation estimated from twin/family studies.[39] For ADHD specifically, 37% of the genetic variation was

estimated to be due to the effects of SNPs included on GWAS platforms (28% of the liability to ADHD).[39] Yet, the only other study that estimated chip heritability for ADHD with GCTA found an estimate of zero for continuous measures of ADHD-like behaviors.[52] Although this discrepancy clearly calls for further investigation, one other study that used a different method supports the presence of a relevant aggregate genetic signal in ADHD GWA meta-analysis results; ADHD polygenic risk scores that summarize the genetic risk of the individual based on the results of the 2010 PGC meta-analysis were shown to be predictive of ADHD case status in an independent sample.[40, 53] While more research is needed to validate these results, altogether, these findings suggest that common variants tagged by GWAS platforms play an important role in ADHD, but stay undetected in current GWAS projects due to limited sample size. Even though the effect sizes of individual variants will be small, if detected, they can lead to a better understanding of the biological basis of ADHD.

Environmental risk factors for AP and ADHD. Many studies have focused on genetic risk factors for ADHD, but this should not lead to the neglect of possible environmental contributions to disease risk, as a considerable part of the variation can be attributed to the environment. A few risk factors have been consistently associated with ADHD in multiple studies of considerable sample size, namely prematurity, low birth weight, lead exposure and extreme early adversity, as noted in reviews by e.g. Thapar et al. and Nigg.[54, 55] Risk factors that have been reported before but for which the evidence is less convincing consist of prenatal factors (maternal stress, smoking and alcohol use during pregnancy), environmental toxins (organophosphate pesticides, polychlorinated biphenyls), dietary factors (deficiencies of zinc, magnesium, polyunsaturated fatty acids, surpluses of sugar and artificial food colorings and low/high IgG foods) and psychosocial adversity (family adversity and low income and conflict/parent-child hostility), as summarized in a review by Thapar et al. (summary taken from Table 2 in this review).[55] A major issue with many of these environmental factors is that, even when an association has been confirmed, it remains to be established whether the observed association is due to a causal effect of the risk factor on ADHD, or whether the association is due to a third factor that influences both the predictor and the outcome, e.g. the genetic make-up of the child or parent.

Gene-environment interplay. Although many studies assess the individual contribution of genetic or environmental risk factors, these two sources of variance can correlate or interact with each

other. Gene-environment correlation refers to the situation in which the genes that influence a trait also influence the chances of exposure to a particular environment that influences the trait. Gene-environment interaction occurs when the expression of genetic factors depends on the environment, or the other way around: a particular environmental risk factor is only a risk factor for subjects with a particular genotype. Twin models can test for the interaction of latent genetic variation and a measured environmental risk factor, but these models have not often been applied to ADHD.[56, 57] A larger portion of the gene-environment interaction literature has been dedicated to measured environments in combination with measured variants in candidate genes. Following two seminal papers by Caspi et al.[58, 59] on violent behavior and on depression, these studies have often led to appealing findings, for example, one study found that breastfeeding has a positive effect on IQ, but only in children carrying at least one major allele of a polymorphism in the FADS2 gene, which is involved in fatty acid metabolism.[60] As with the candidate gene association literature, candidate gene-environment interaction studies have been criticized for a lack of replication and concern of unreported multiple testing in the context of low power in small samples.[61] With regard to ADHD specifically, some evidence exists for an interaction of serotonergic and dopaminergic genotypes with psychosocial factors, as summarized in a review on measured gene by environment interaction studies by Nigg et al.[56] However, most of the studies included in this review are of small sample size, with varying definitions so that these results should be interpreted with caution. Yet the hypothetical appeal of the gene-environment framework is untouched and the debate on the appropriate evaluation of these interactions is ongoing, discussing among other things, appropriate statistical models and accurate measurement of environmental exposures.[62, 63] Most likely, gene-environment interplay studies will advance further when larger GWA studies have identified replicable genetic variants for ADHD.

Early indicators of vulnerability. ADHD is a diagnosis typically made at school-age, yet more and more studies have start to focus on hyperactive and inattentive behavior at preschool age.[64-66] Besides these behaviors that mimic later ADHD behaviors closely, there are also other factors early in life that are associated with ADHD, more likely due to a shared vulnerability than through a causal relation. Among these factors are excessive crying behavior and difficult temperament, and a delay in motor development.[67, 68] These factors are important as they may shed light on the etiology of ADHD, and can be helpful in identifying children at risk for ADHD that might benefit from early intervention.

In summary, twin, adoption and family studies have previously established the large contribution of genetic factors to the etiology of Attention Problems and ADHD. Current advances in genetic research allow for the assessment of the effects of large number of individual genetic variants. These new methods can help to answer questions on the overall contribution of genetic factors to a trait, the presence of gene-environment interaction, the stability of genetic effects over time and the genetic overlap across disorders. In addition, the contribution of rare versus common variants can be assessed and specific genetic variants can be identified that play a role in the etiology of a trait. Other epidemiological studies have established the association of AP/ADHD with specific environmental factors, but more research is needed to confirm these associations and determine whether a causal relationship underlies these associations. This thesis aims to contribute to the clarification of the etiology of ADHD by studying specific genetic and environmental risk factors for AP and ADHD.

Outline of the current thesis. The first part of this thesis focuses on early predictors and risk factors of AP. In chapter two, the heritability of crying behavior at age two is investigated, as well as its predictive value for internalizing, externalizing behavior and attention problems at age seven. The third chapter focuses on the association between low birth weight and AP; a causal model is tested by assessing whether a similar effect on AP is observed in unrelated pairs and monozygotic and dizygotic twin pairs discordant for birth weight. The fourth chapter focuses on breastfeeding, and analyzes its association with IQ, educational attainment and attention problems. Further, it is examined whether the effects of breastfeeding are moderated by polymorphism in the FADS2 gene. The second part of the thesis describes several analyses that have been performed to identify genetic variants for ADHD and other pediatric phenotypes, and to assess the contribution of common SNPs. In the fifth chapter, polygenic risk scores based on a GWA of clinical ADHD cases are tested for their predictive value of mother and teacher ratings of attention problems in an independent general-population sample of preschool and school-age children. Chapter six reports on the results of a GWA meta-analysis on continuous measures of ADHD in a consortium of population-based cohorts. In chapter seven, an overview is presented of several contributions to GWA consortia projects that were based on data from the Netherlands Twin Register. These include GWA studies on a wide range of childhood phenotypes such as birth weight, eczema, height and BMI in childhood, age at walking and pubertal development. In

chapter eight the results from this thesis are summarized and an effort is made to discuss the findings in the context of the available literature, and provide directions for further research.

References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Arlington, VA: American Psychiatric Publishing; 2013.

2. Willcutt EG. The prevalence of DSM-IV attention-deficit/hyperactivity disorder: a meta-analytic review. Neurotherapeutics. 2012;9:490-499.

3. Bernfort L, Nordfeldt S, Persson J. ADHD from a socio-economic perspective. Acta Paediatr. 2008;97:239-245.

4. Matza LS, Paramore C, Prasad M. A review of the economic burden of ADHD. Cost Eff Resour Alloc. 2005;3:5.

5. Schachter HM, Pham B, King J, Langford S, Moher D. How efficacious and safe is short-acting methylphenidate for the treatment of attention-deficit disorder in children and adolescents? A meta-analysis. CMAJ. 2001;165:1475-1488.

6. Sonuga-Barke EJ, Brandeis D, Cortese S, et al. Nonpharmacological interventions for ADHD: systematic review and meta-analyses of randomized controlled trials of dietary and psychological treatments. Am J Psychiatry. 2013;170:275-289.

7. Danckaerts M, Sonuga-Barke EJ, Banaschewski T, et al. The quality of life of children with attention deficit/hyperactivity disorder: a systematic review. Eur Child Adolesc Psychiatry. 2010;19:83-105.

8. Willcutt EG, Nigg JT, Pennington BF, et al. Validity of DSM-IV attention deficit/hyperactivity disorder symptom dimensions and subtypes. J Abnorm Psychol. 2012;121:991-1010.

9. de Graaf R, Kessler RC, Fayyad J, et al. The prevalence and effects of adult attention-deficit/hyperactivity disorder (ADHD) on the performance of workers: results from the WHO World Mental Health Survey Initiative. Occup Environ Med. 2008;65:835-842.

10. Mannuzza S, Klein RG, Bessler A, Malloy P, LaPadula M. Adult outcome of hyperactive boys. Educational achievement, occupational rank, and psychiatric status. Arch Gen Psychiatry. 1993;50:565-576.

11. Bradshaw JL, Sheppard DM. The neurodevelopmental frontostriatal disorders: evolutionary adaptiveness and anomalous lateralization. Brain Lang. 2000;73:297-320.

12. Jensen PS, Mrazek D, Knapp PK, et al. Evolution and revolution in child psychiatry: ADHD as a disorder of adaptation. J Am Acad Child Adolesc Psychiatry. 1997;36:1672-1679; discussion 1679-1681.

13. Verheul I, Block J, Burmeister-Lamp K, Thurik R, Tiemeier H, Turturea R. ADHD-like behavior and entrepreneurial intentions. 2013. http://repub.eur.nl/pub/37266.

14. White HA, Shah P. Creative style and achievement in adults with attention-deficit/hyperactivity disorder. Pers Indiv Differ. 2011;50:673-677.

15. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed, text revision. Washington, DC: American Psychiatric Association; 2000.

16. World Health Organisation. ICD-10 Classifications of Mental and Behavioural Disorder: Clinical Descriptions and Diagnostic Guidelines. Geneva: World Health Organisation; 1992.

17. Hudziak JJ, Achenbach TM, Althoff RR, Pine DS. A dimensional approach to developmental psychopathology. Int J Methods Psychiatr Res. 2007;16 Suppl 1:S16-23.

18. Swanson JM, Schuck S, Mann M, et al. Categorical and dimensional definitions and evaluations of symptoms of ADHD: The SNAP and SWAN Rating Scales. 2006. Retrieved 1 November 2013 from http://www.ADHD.net/SNAP_SWAN.pdf.

19. Achenbach TM, Rescorla LA. Manual for the ASEBA Preschool forms & Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families; 2000.

20. Achenbach TM, Rescorla LA. Manual for the ASEBA School-Age Forms & Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families; 2001.

21. Goodman R. The Strengths and Difficulties Questionnaire: a research note. J Child Psychol Psychiatry. 1997;38:581-586.

22. Derks EM, Hudziak JJ, Dolan CV, Ferdinand RF, Boomsma DI. The relations between DISC-IV DSM diagnoses of ADHD and multi-informant CBCL-AP syndrome scores. Compr Psychiatry. 2006;47:116-122.

23. Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. Br J Psychiatry. 2000;177:534-539.

24. Hudziak JJ, Heath AC, Madden PF, et al. Latent class and factor analysis of DSM-IV ADHD: a twin study of female adolescents. J Am Acad Child Adolesc Psychiatry. 1998;37:848-857.

25. Lubke GH, Hudziak JJ, Derks EM, van Bijsterveldt TC, Boomsma DI. Maternal ratings of attention problems in ADHD: evidence for the existence of a continuum. J Am Acad Child Adolesc Psychiatry. 2009;48:1085-1093.

26. Lubke GH, Muthen B, Moilanen IK, et al. Subtypes versus severity differences in attention-deficit/hyperactivity disorder in the Northern Finnish Birth Cohort. J Am Acad Child Adolesc Psychiatry. 2007;46:1584-1593.

27. Neuman RJ, Todd RD, Heath AC, et al. Evaluation of ADHD typology in three contrasting samples: a latent class approach. J Am Acad Child Adolesc Psychiatry. 1999;38:25-33.

28. Todd RD, Rasmussen ER, Neuman RJ, et al. Familiality and heritability of subtypes of attention deficit hyperactivity disorder in a population sample of adolescent female twins. Am J Psychiatry. 2001;158:1891-1898.

29. Eaves LJ, Silberg JL, Meyer JM, et al. Genetics and developmental psychopathology: 2. The main effects of genes and environment on behavioral problems in the Virginia Twin Study of Adolescent Behavioral Development. J Child Psychol Psychiatry. 1997;38:965-980.

30. Faraone SV, Perlis RH, Doyle AE, et al. Molecular genetics of attention-deficit/hyperactivity disorder. Biol Psychiatry. 2005;57:1313-1323.

31. Gjone H, Stevenson J, Sundet JM. Genetic influence on parent-reported attention-related problems in a Norwegian general population twin sample. J Am Acad Child Adolesc Psychiatry. 1996;35:588-596; discussion 596-588.

32. Hudziak JJ, Rudiger LP, Neale MC, Heath AC, Todd RD. A twin study of inattentive, aggressive, and anxious/depressed behaviors. J Am Acad Child Adolesc Psychiatry. 2000;39:469-476.

33. Rietveld MJ, Hudziak JJ, Bartels M, van Beijsterveldt CE, Boomsma DI. Heritability of attention problems in children: I. cross-sectional results from a study of twins, age 3-12 years. Am J Med Genet B Neuropsychiatr Genet. 2003;117B:102-113.

34. Sherman DK, McGue MK, Iacono WG. Twin concordance for attention deficit hyperactivity disorder: a comparison of teachers' and mothers' reports. Am J Psychiatry. 1997;154:532-535.

35. Sprich S, Biederman J, Crawford MH, Mundy E, Faraone SV. Adoptive and biological families of children and adolescents with ADHD. J Am Acad Child Adolesc Psychiatry. 2000;39:1432-1437.

36. Gizer IR, Ficks C, Waldman ID. Candidate gene studies of ADHD: a meta-analytic review. Hum Genet. 2009;126:51-90.

37. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. Science. 2005;308:385-389.

38. A Catalog of Published Genome-Wide Association Studies. 2013. www.genome.gov/gwastudies. Accessed 23 September 2013.

39. Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee SH, Ripke S, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet. 2013;45:984-994.

40. Neale BM, Medland SE, Ripke S, et al. Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry. 2010;49:884-897.

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

42. Maher B. Personal genomes: The case of the missing heritability. Nature. 2008;456:18-21.

43. Kan KJ, van Beijsterveldt CEM, Bartels M, Boomsma DI. Assessing genetic influences on behavior: Informant and context dependency as illustrated by the analysis of Attention Problems. Under review.

44. Antshel KM, Fremont W, Roizen NJ, et al. ADHD, major depressive disorder, and simple phobias are prevalent psychiatric conditions in youth with velocardiofacial syndrome. J Am Acad Child Adolesc Psychiatry. 2006;45:596-603.

45. Ehli EA, Abdellaoui A, Hu Y, et al. De novo and inherited CNVs in MZ twin pairs selected for discordance and concordance on Attention Problems. Eur J Hum Genet. 2012;20:1037-1043.

46. Williams NM, Zaharieva I, Martin A, et al. Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. Lancet. 2010;376:1401-1408.

47. Yang J, Manolio TA, Pasquale LR, et al. Genome partitioning of genetic variation for complex traits using common SNPs. Nat Genet. 2011;43:519-525.

48. So HC, Li M, Sham PC. Uncovering the total heritability explained by all true susceptibility variants in a genome-wide association study. Genet Epidemiol. 2011;35:447-456.

49. Davies G, Tenesa A, Payton A, et al. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. Mol Psychiatry. 2011;16:996-1005.

50. Lubke GH, Hottenga JJ, Walters R, et al. Estimating the genetic variance of major depressive disorder due to all single nucleotide polymorphisms. Biol Psychiatry. 2012;72:707-709.

51. Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42:565-569.

52. Trzaskowski M, Dale PS, Plomin R. No genetic influence for childhood behavior problems from DNA analysis. J Am Acad Child Adolesc Psychiatry. 2013;52:1048-1056.e3.

53. Hamshere ML, Langley K, Martin J, et al. High loading of polygenic risk for ADHD in children with comorbid aggression. Am J Psychiatry. 2013;170:909-916.

54. Nigg JT. What causes ADHD? New York: Guilford Press; 2006.

55. Thapar A, Cooper M, Eyre O, Langley K. Practitioner review: what have we learnt about the causes of ADHD? Journal of child psychology and psychiatry, and allied disciplines. 2013;54:3-16.

56. Nigg J, Nikolas M, Burt SA. Measured gene-by-environment interaction in relation to attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry. 2010;49:863-873.

57. Purcell S. Variance components models for gene-environment interaction in twin analysis. Twin Res. 2002;5:554-571.

58. Caspi A, McClay J, Moffitt TE, et al. Role of genotype in the cycle of violence in maltreated children. Science. 2002;297:851-854.

59. Caspi A, Sugden K, Moffitt TE, et al. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. Science. 2003;301:386-389.

60. Caspi A, Williams B, Kim-Cohen J, et al. Moderation of breastfeeding effects on the IQ by genetic variation in fatty acid metabolism. Proceedings of the National Academy of Sciences of the United States of America. 2007;104:18860-18865.

61. Duncan LE, Keller MC. A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. Am J Psychiatry. 2011;168:1041-1049.

62. Belsky J, Pluess M, Widaman KF. Confirmatory and competitive evaluation of alternative gene-environment interaction hypotheses. J Child Psychol Psychiatry. 2013;54:1135-1143.

63. Caspi A, Hariri AR, Holmes A, Uher R, Moffitt TE. Genetic sensitivity to the environment: the case of the serotonin transporter gene and its implications for studying complex diseases and traits. Am J Psychiatry. 2010;167:509-527.

64. Greenhill LL, Posner K, Vaughan BS, Kratochvil CJ. Attention deficit hyperactivity disorder in preschool children. Child Adolesc Psychiatr Clin N Am. 2008;17:347-366, ix.

65. Lahey BB, Pelham WE, Loney J, et al. Three-year predictive validity of DSM-IV attention deficit hyperactivity disorder in children diagnosed at 4-6 years of age. Am J Psychiatry. 2004;161:2014-2020.

66. Riddle MA, Yershova K, Lazzaretto D, et al. The Preschool Attention-Deficit/Hyperactivity Disorder Treatment Study (PATS) 6-year follow-up. J Am Acad Child Adolesc Psychiatry. 2013;52:264-278.e2.

67. Hemmi MH, Wolke D, Schneider S. Associations between problems with crying, sleeping and/or feeding in infancy and long-term behavioural outcomes in childhood: a meta-analysis. Arch Dis Child. 2011;96:622-629.

68. Perna R, Loughan A. Early developmental delays: neuropsychological sequelae and subsequent diagnoses. Applied neuropsychology. Child. 2012;1:57-62.

Part I

Early predictors and risk factors for Attention Problems

Chapter 2

Crying without a cause and being easily upset in two-year olds: heritability and predictive power of behavioral problems

Groen-Blokhuis MM, Middeldorp CM, van Beijsterveldt CEM, Boomsma DI. Crying without a cause and being easily upset in two-year-olds: heritability and predictive power of behavioral problems. Twin Res Hum Genet 2011;14:393-400.

Abstract

In order to estimate the influence of genetic and environmental factors on “crying without a cause” and “being easily upset” in 2-year old children, a large twin study was carried out. Prospective data were available for ~18,000 two year old twin pairs from the Netherlands Twin Register. A bivariate genetic analysis was performed using structural equation modeling in the Mx software package. The influence of maternal personality characteristics and demographic and lifestyle factors was tested to identify specific risk factors that may underlie the shared environment of twins. Furthermore, it was tested whether crying without a cause and being easily upset were predictive of later internalizing, externalizing and attention problems.

Crying without a cause yielded a heritability estimate of 60% in boys and girls. For easily upset, the heritability was estimated at 43% in boys and 31% in girls. The variance explained by shared environment varied between 35% and 63%. The correlation between crying without a cause and easily upset ($r = 0.36$) was explained both by genetic and shared environmental factors. Birth cohort, gestational age, socioeconomic status, parental age, parental smoking behavior and alcohol use during pregnancy did not explain the shared environmental component. Neuroticism of the mother explained a small proportion of the additive genetic, but not of the shared environmental effects for easily upset. Crying without a cause and being easily upset at age 2 were predictive of internalizing, externalizing and attention problems at age 7, with effect sizes of 0.28-0.42.

A large influence of shared environmental factors on crying without a cause and easily upset was detected. Although these effects could be specific to these items, we could not explain them by personality characteristics of the mother or by demographic and lifestyle factors and we recognize that these effects may reflect other maternal characteristics. A substantial influence of genetic factors was found for the two items, which are predictive of later behavioral problems.

Introduction

A significant number of children cry persistently in the first 3 to 6 months of life. Prevalence rates vary between 0.3% and 7.7% depending on the definition, but are usually reported to be around 5%.[1] Children displaying high levels of persistent crying and distress are a source of concern to their parents and a frequent reason to consult a family doctor or pediatrician.[2] After the first months of life, levels of crying decrease and crying behavior becomes a more stable characteristic of the child. Children who frequently fuss or cry are thought of as having a difficult temperament. Difficult temperament shows significant continuity over time and is associated with internalizing and externalizing problem behavior and attention problems later in life.[3-7]

A recent review of twin and adoption studies reported that the heritability of temperament varied between 23% and 60%.[8] Studies investigating specifically the Emotionality and Irritability/Anger subscales with items describing crying behavior and proneness to distress yielded heritability estimates of 42-72%.[9-11] The remaining part of the variance is mostly explained by unique environmental factors. Shared environment accounts for a small proportion of the variance in most temperament studies, and its influence is often considered negligible.[8] The studies included in the review varied regarding sample size from 50 to 800 individuals.

Most of these studies investigated the heritability of temperament dimensions by means of a sum score of a set of items that are thought to measure an underlying construct like negative emotionality. However, studies on personality have shown that genetic and environmental effects can be facet or item specific; these effects are not picked up when broad dimensions are analyzed.[12, 13] A methodological study showed that variance components estimates based on sum scores can differ substantially from the variance components of the underlying latent trait and proposed to use multivariate genetic analysis at the item or symptom level instead of sum scores.[14] The literature on focal aspects of temperament is rather limited, but one study that estimated the influence of genetic and environmental factors on focal aspects and broader dimensions of temperament, found some genetic or shared environmental influences to be facet specific.[10]

The present study focuses on two items indicating two aspects of negative emotionality, namely crying without a cause (CWC) and being easily upset (EU). Data were available for more than 18,000 two year old Dutch twin pairs. The aim of the study was to estimate the heritability of these two specific behaviors and to establish the contribution of the shared and unique environment to these traits. The twins and their parents participate in a longitudinal survey study of the Netherlands Twin Register and data on environmental factors and maternal characteristics were tested for their relation to CWC and EU in an attempt to assess the shared environment of the twins. In addition, it was investigated in a subsample of 8,994 twins whether CWC and EU at age two were predictive of internalizing, externalizing and attention problems at age 7.

Materials and methods

Participants. The Netherlands Twin Register (NTR) is a population based register that was established at the VU University Amsterdam in 1986. New born twins are enrolled in longitudinal survey projects.[15] Parents receive questionnaires by mail until the twins are 12 years old.

Since 1988, two items indicating crying without a cause and being easily upset have been included in the survey that was sent out to all mothers of two-year-old twins. Over 18,000 parents filled in and returned this questionnaire with a response rate of 83.4%. In later versions of the questionnaire respondents were asked to indicate their relationship to the twins; the respondent was the mother of the children in 93.8% of the cases, the father of the children in 5.1% of the cases and had a relationship specified as 'other' in 1.1% of the cases. The number of twin pairs was almost equally distributed over cohorts 1986 – 2004.

Procedures and Instruments

Zygosity. If available, DNA or blood group testing was taken as the conclusive result for zygosity determination. Zygosity for the remaining same-sex twin pairs was determined by a set of questionnaire items filled in by the parents of the twins through ages 3-12. This instrument correctly determines zygosity in 93% of same-sex twin pairs.[16] In 11.3% of the cases, zygosity was determined by a single item in the questionnaire that was sent out at age 2, that indicates how much the children look alike. This question gives a correct determination of zygosity in 92% of the cases. For 16 twin pairs, zygosity determination was not available and they were excluded from the study.

Measures. The survey sent out at age 2 included two items that describe facets of negative emotionality. The first question ‘Did the children cry without a clear cause?’ could be answered with ‘rarely’, ‘sometimes’ or ‘often’. The second question ‘Are the children upset for a long time if the usual course of events is disrupted?’ could be answered with ‘yes’, ‘a bit’ or ‘never/hardly ever’. Both items were rated for each child separately.

Data on parental age, maternal tobacco and alcohol consumption during pregnancy, paternal smoking behavior during pregnancy and gestational age were based on the first survey sent out

after registration with the NTR. For the present study, data on maternal tobacco use during pregnancy were coded in 2 categories; 'yes' and 'no', data on paternal tobacco use during pregnancy were coded in 4 categories; 'no', 'yes, cigars/pipe', 'yes, < 10 cigarettes per day', 'yes, > 10 cigarettes per day' and data on maternal alcohol use during pregnancy were coded in 3 answer categories; 'no', 'yes, < 1 glass per week' and 'yes, > 1 glass per week'. Data on demographic and life-style factors and parental personality traits were available from several surveys. Socioeconomic status (SES) was based on a full description of parental occupation at age 3 in two third of the families and coded according to the standard classification of occupations.[17] For the remaining families, parental occupational status was based on Goldthorpe's class categories (EPG) combined with information on parental educational level at age 3.[18] Complete data on demographic and life-style factors were available for 13,065 twin pairs. Incomplete data were most often due to a lack of data on SES as a result of drop out from the study at age 3. Data on Neuroticism, Extraversion, Openness, Altruism and Conscientiousness were collected using 60 items of the NEO Five Factor Inventory (NEO-FFI),[19] which were included in a survey in 2009-2010 sent out to parents and twins aged 18 years and older registered within the NTR. Personality data were available for 1,040 mothers of twins.

An age appropriate version of the Child and Behavior Checklist (CBCL 4-18) was included in the survey that is sent out at age 7. This checklist consists of 120 items that describe several behavioral problems. Parents are asked to rate the behavior of their child during the preceding 6 months on a three point scale, 0='not true', 1='somewhat or sometimes true', 2='very true or often true'. The attention problems subscale describes both hyperactive and inattentive behaviours. The internalizing scale consists of the anxious/depressed, somatic complaints, and withdrawn subscales. The externalizing scale consists of the aggressive and rule-breaking behavior subscales.[20, 21] Data on CWC and EU at age 2 and internalizing, externalizing and attention problems at age 7 were available for 8,994 twin pairs.

Statistical analysis. Monozygotic (MZ) twins are genetically nearly identical while dizygotic (DZ) twins share on average 50% of their segregating genes. A higher resemblance of MZ twins compared to DZ twins for a particular trait is therefore likely to be due to the influence of genetic factors. Environmental factors that are shared between the twins will make MZ and DZ twins resemble each other, whereas unique environmental experiences will cause both MZ and DZ

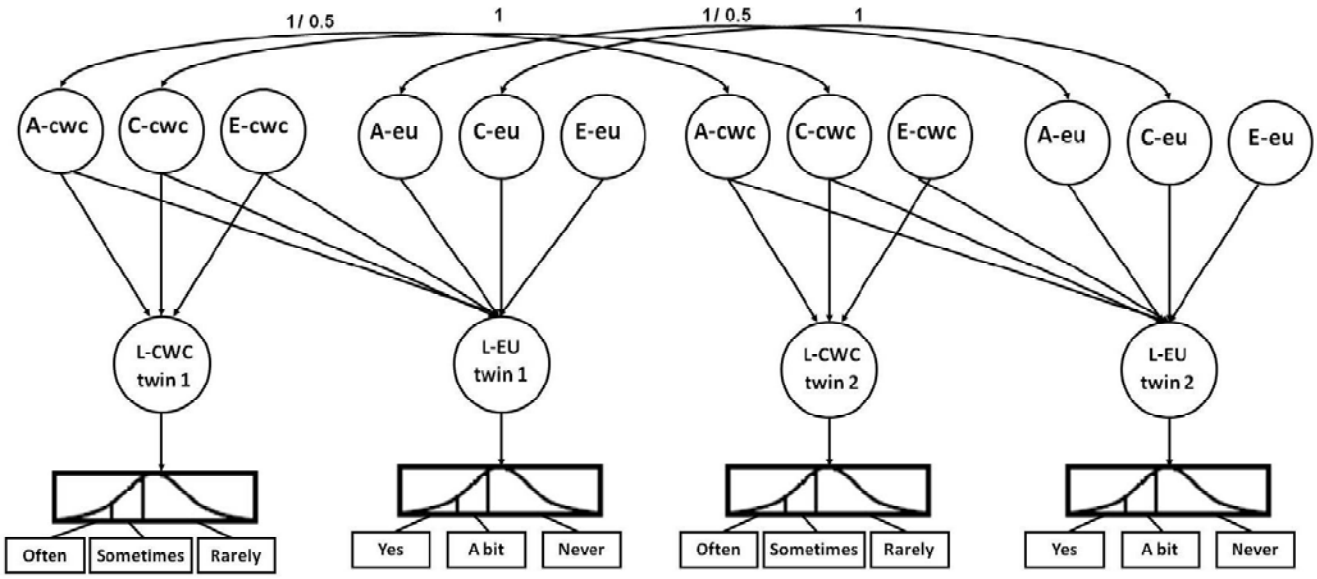
twins to differ from each other. When phenotypic data are available on MZ and DZ twin pairs, the total variance of the trait can be decomposed into variance due to genetic factors (A), common environment (C) shared by children from the same family and unique environment (E). When bivariate data are analyzed, it is possible to detect the causes of covariance between two traits; that is the proportion of covariance that is due to additive genetic factors, shared environment or unique environment. For instance, using the same logic as before, when the correlation across-traits, across-twins is higher in MZ than in DZ twins (e.g. CWC in one twin and EU in the co-twin), it can be concluded that genetic factors common to both traits partly explain the covariance between the traits. For a more comprehensive summary of the methods in twin research, including tests for sex x genotype interaction we refer to Martin et al. and Boomsma et al.[22, 23]

Liability threshold model. The two categorical outcome variables were analyzed using a liability threshold model. This model assumes that an unobserved liability underlies the measured categories of CWC and EU. The liability can be influenced by an individual's genetic makeup and exposure to environmental influences and is normally distributed with a mean value of 0 and a variance of 1. If a threshold on the liability scale is passed, children enter the next category of the ordinal scale. The thresholds are calculated as the number of standard deviations away from the mean. By definition, the area under the curve corresponds to the probability to be in a certain category (Figure 1). The resemblance between the first and second born of a twin pair is estimated by the correlation for the liability scale, called a polychoric correlation. Polychoric twin correlations were estimated using the software package Mx in a so called saturated model, in which the correlations and thresholds are estimated unconstrained over the different zygosity / sex groups. Then, thresholds were constrained to be equal over the different groups to test for effects of sex and zygosity. The fit of these models was compared with the saturated model. Analyses were performed on raw data using the full information maximum likelihood (FIML) method. The fit of the different models was compared by the log-likelihood ratio test (LRT). The difference in minus two times the log-likelihood (-2LL) between two models has a χ^2 distribution with the degrees of freedom (*df*) equaling the difference in *df* between the two models. As we used a very large dataset, a p-value < 0.01 was taken as significant.

Next, a bivariate genetic model (Figure 1) was evaluated. Here, the variance in liability was decomposed using a model with three latent variables: additive genetic factors (A), shared environment (C) and unique environment (E). The variance in liability explained by A, C and E is calculated by squaring the factor loadings, which sum to 1. MZ twins correlate 1 for A and DZ twins correlate on average 0.5. By definition, C correlates 1 in all twin pairs, whereas E correlates 0. In a bivariate model, the genetic factors that influence CWC also load on the liability to EU, and the same is the case for the environmental factors. The Mx software package evaluates the fit of different values of the parameter estimates (i.e. factor loadings and thresholds) and provides estimates that offer the best fit of the model given the data.

Figure 1: Path diagram of the bivariate model with the latent factors A, C and E and their influence on the liability to CWC and EU in twin 1 and 2 as modeled in the liability threshold model.

A=additive genetic factors, C=shared environment, E=unique environment, L=liability to the trait, CWC=crying without a cause, EU=easily upset



A series of models was tested. Firstly, an ACE model was fitted to the data that allowed estimates to differ in boys and girls. Then, the loadings on the A, C and E factors were constrained to be equal between sexes. Next, the AE and CE model were tested against the ACE model. Finally, we aimed to identify specific risk factors that could underlie the shared environment. Associations were tested between CWC and EU and birth cohort, socioeconomic status, parental age, maternal tobacco and alcohol consumption during pregnancy, paternal smoking behavior during pregnancy and gestational age. The thresholds of CWC and EU were regressed on these variables by including them in the bivariate model as fixed effects on the thresholds. These predictors were then dropped one by one and tested for a difference in model fit. Following the same procedure, it was tested if personality characteristics measured in a subsample of the mothers influenced the ratings of the children's behavior.

To test whether CWC and EU were predictive of internalizing, externalizing and attention problems at age 7, the children who were crying without a cause 'often' were compared to the children who were crying without a cause 'rarely' and 'sometimes', and the children who were easily upset 'yes', were compared to the children who were easily upset 'a bit' or 'never/hardly ever'. Data from two individuals within a twin pair are not independent, therefore effect sizes were calculated for first born twins only. Effect sizes were calculated in terms of Cohen's d: the difference between two means divided by the pooled SD's for those means.[24]

Results

Crying Without a Cause was reported to occur rarely in 64.2%, sometimes in 27.2% and often in 8.6% of the cases. Children were Easily Upset 'yes' in 5.9%, 'a bit' in 34.8% and 'never/hardly ever' in 59.3% of the cases. These frequencies are in line with general reports on the frequency of crying behavior.[1]

In the saturated model, thresholds could be constrained to be equal over all groups without a significant worsening of the fit ($p=0.063$), except for the threshold of girls who were part of an opposite-sex (DOS) twin pair ($p < 0.001$). Thus, there were no overall sex differences in item responses, but girls with a male co-twin were consistently rated as less easily upset and less often crying without a cause than girls with a female co-twin. The mirror effect was not observed, that is, boys with a female co-twin do not differ from boys with a male co-twin. The phenotypic correlation for CWC and EU was 0.36 in boys and girls.

The polychoric twin correlations and the cross-trait cross-twin correlations (the correlation between CWC in one twin and EU in the co-twin) were estimated in the saturated model (Table 1). The within-trait and the cross-trait twin correlations were higher in MZ than in DZ twins, providing evidence for genetic influences on CWC and EU and on the covariation between the traits. However, MZ correlations were never twice as high as DZ correlations, indicating a role for shared environmental influences. Next, the ACE model was evaluated. The factor loadings were significantly different for boys and girls for EU, but not for CWC. Therefore, it was tested separately for males and females whether the influence of A, C and E on EU and the cross loadings from CWC on EU could be dropped from the model. The influence of A and C could not be dropped from the model without a significant worse fit, both for CWC and EU. None of the cross loadings could be dropped from the model (Table 2).

Table 1: Polychoric twin correlations based on maximum likelihood estimates of a liability model with 3 answer categories.

	Number of twin pairs	CWC twin correlation	EU twin correlation	Cross-trait cross-twin correlation

MZM	2884	0.95	0.94	0.33
DZM	3165	0.64	0.73	0.24
MZF	3228	0.96	0.94	0.34
DZF	2861	0.64	0.78	0.25
DOS	6153	0.66	0.75	0.24

CWC=Crying without a cause, EU=Easily upset, MZM=monozygotic males, DZM=dizygotic males, MZF=monozygotic females, DZF=dizygotic females, DOS=dizygotic opposite-sex twin pairs

Table 2: Model fitting results for the bivariate model of Crying without a cause (CWC) and Easily Upset (EU). The best model is specified in bold font. As the factor loadings of EU showed significant sex differences, it was tested separately for males (M) and females (F) whether the influence of A, C, E on EU and the cross loadings of CWC on EU could be dropped from the model.

Model	-2 times Likelihood	Df	Comparison	X^2 difference	Difference in df	p-value
1. ACE model with sex differences on the path loadings	101009.297	72762				
2. ACE model no sex differences	101034.315	72769	Model 1	25.018	7	< 0.001
3. ACE model no sex differences CWC path loadings	101018.424	72764	Model 1	9.127	2	0.010
4. AE-ACE model	101367.745	72765	Model 3	349.321	1	< 0.001
5. ACE-AE model (M) ACE-AE model (F)	101472.907 101545.533	72766 72766	Model 3 Model 3	454.483 527.109	2 2	< 0.001 < 0.001
6. CE-ACE-model	102629.857	72765	Model 3	1611.433	1	< 0.001
7. ACE-CE-model (M) ACE-CE-model (F)	102130.111 102128.917	72766 72766	Model 3 Model 3	1111.687 1110.493	2 2	< 0.001 < 0.001
8. Drop A cross loading (M) Drop A cross loading (F)	101152.551 101135.759	72765 72765	Model 3 Model 3	134.127 117.335	1 1	< 0.001 < 0.001
9. Drop C cross loading (M)	101098.258	72765	Model 3	79.834	1	< 0.001

Drop C cross loading (F)	101098.317	72765	Model 3	79.893	1	< 0.001
10. Drop E cross loading (M)	101106.749	72765	Model 3	88.325	1	< 0.001
Drop E cross loading (F)	101073.005	72765	Model 3	54.581	1	< 0.001

CWC=crying without a cause, EU=easily upset, A=additive genetic factors, C=shared environment, E=unique environment, M=males, F=females

Parameter estimates are provided in Table 3. The correlation between the additive genetic factors influencing CWC and the genetic factors influencing EU was 0.36 in boys and 0.41 in girls. The correlation between shared environment influencing CWC and the shared environment influencing EU was 0.36 in boys and 0.34 in girls.

Table 3: Percentage of variance and covariance explained by additive genetic influences (A), common environment shared by twins (C) and unique environment (E) as estimated in the best model specified in Table 2.

Variance component	A (%)	C (%)	E (%)
CWC	60.4	34.7	4.9
EU (M)	42.5	51.7	5.8
EU(F)	30.5	63.1	6.4
Covariance CWC-EU explained (M)	49.8	41.6	8.6
Covariance CWC-EU explained (F)	48.4	43.8	7.8

CWC=crying without a cause, EU=easily upset, M=males, F=females

As a large influence of C was found, several demographic and lifestyle factors were tested for their influence on CWC and EU in a subsample of twin families (n=13,065), the results are shown in Table 4. Gestational age was the only predictor that could not be dropped from the model without resulting in a significant worse fit. However, the differences in A, C and E loadings in the model including all versus no predictors were very small. Moreover, the predictors influenced the factor loadings of A on CWC and EU rather than the factor loadings of C. Finally, maternal NEO scores were examined as predictors of CWC and EU (n=1,040), the results are shown in Table 5. Only Neuroticism of the mother significantly predicted EU, but not

CWC. Including all personality scales as predictors in the model reduced the proportion of variance due to genetic factors with 2%. Again, the influence of shared environment was unchanged.

Table 4: Results of the regression of the thresholds of CWC and EU ratings on several demographic and lifestyle factors.

	CWC			**EU**	
Model	Chi square	p-value		Chi square	p-value
1. Drop birth cohort	0.049	0.826		0.231	0.631
2. Drop SES	4.211	0.040		2.832	0.092
3. Drop maternal age	0.372	0.542		7.138	0.008
4. Drop paternal age	1.420	0.233		0.020	0.887
5. Drop maternal smoking behavior	0.321	0.571		1.531	0.472
6. Drop paternal smoking behavior	0.226	0.634		0.004	0.951
7. Drop maternal alcohol use	0.030	0.863		1.399	0.237
8. Drop gestational age	27.306	< 0.001		15.343	< 0.001
9. Drop all	41.739	< 0.001		44.813	< 0.001

CWC=Crying without a cause, EU=Easily upset

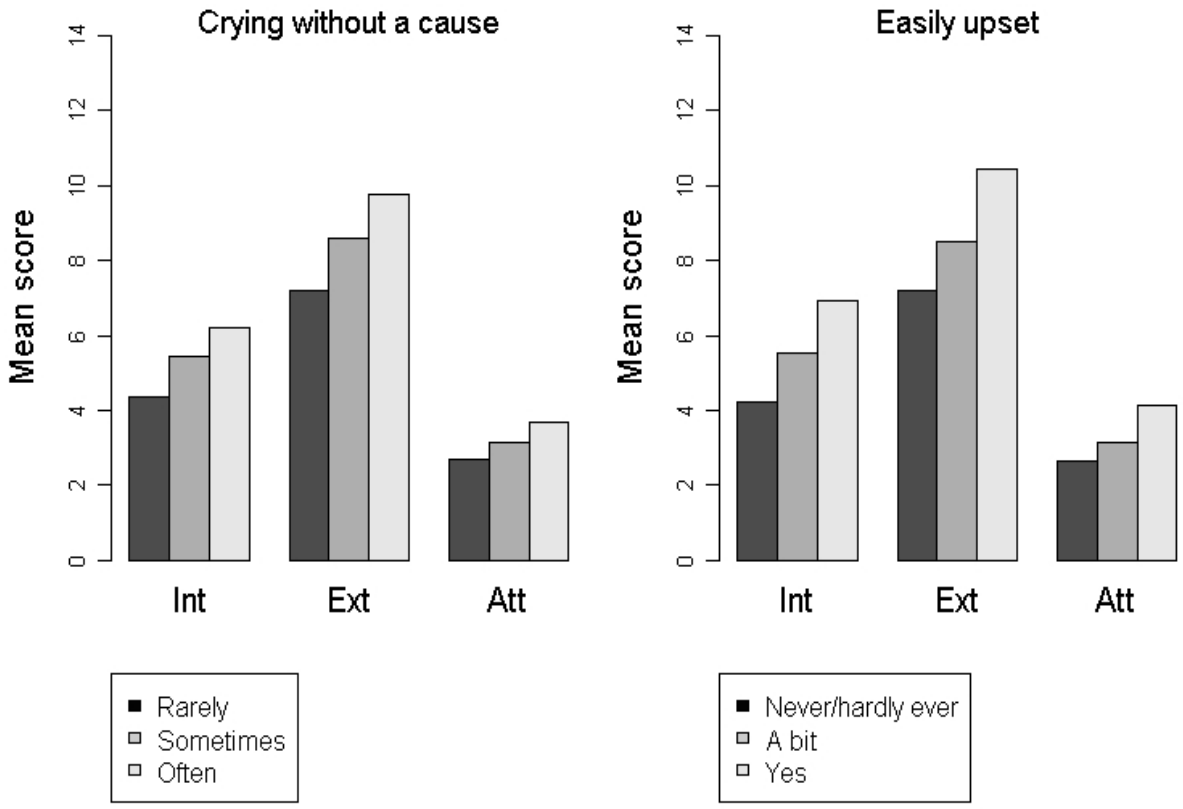
Table 5: Results of the regression of the thresholds of CWC an EU ratings on the NEO scores of the mother

	CWC			**EU**	
Model	Chi square	p-value		Chi square	p-value
1. Drop Neuroticism	3.799	0.051		30.054	< 0.001
2. Drop Extraversion	0.020	0.888		0.030	0.864
3. Drop Openness	0.390	0.532		0.020	0.888
4. Drop Agreeableness	0.020	0.887		0.573	0.449
5. Drop Conscientiousness	0.656	0.418		1.200	0.273
6. Drop all	5.486	0.360		37.405	< 0.001

CWC=Crying without a cause, EU=Easily upset

The CWC and EU items were predictive of internalizing, externalizing and attention problems at age 7: first born twins who were often crying without a cause and/or easily upset scored significantly higher on all three scales, as is depicted in Figure 2. Effect sizes were .30, .29 and .28 for CWC and .42, .38 and .41 for EU respectively. The polychoric correlations were .15, .15, and .13 for CWC and .19, .15 and .15 for EU.

Figure 2: Mean internalizing (Int), externalizing (Ext) and attention (Att) problems scores at age 7 for first born twins per answer category of the items Crying without a cause and Easily upset.



Discussion

In this large sample of two year old twins, the heritability of CWC and EU was estimated between 30% and 60%. Shared environment explained 35-63% of the variance in CWC and EU. The unique environment, which by definition includes measurement error, explained only a small proportion of variance, around 5%. The large estimate of shared environmental influence was not explained by the effect of birth cohort, socioeconomic status, parental age, gestational age, parental smoking behavior and alcohol use during pregnancy, neither by personality characteristics of the mother.

Previous studies on the heritability of crying behavior after the first 3-6 months of life are mostly performed within the temperament framework. The Emotionality and Irritability/Anger subscales showed heritability estimates between 42-72% and negligible influence of shared family environment, both in studies using parental ratings and studies utilizing laboratory observations.[9-11] The most obvious difference between the current and previous studies is that we assessed two specific behaviors with two items on a 3-point scale, whereas other studies utilized extensive temperament questionnaires and analyzed the according dimensions. It is therefore possible that we have detected effects that are unique to these specific behaviors. Multivariate item genetic analyses of temperament questionnaires that are more commonly used should clarify which genetic and shared environmental effects are shared among the items and which are item-specific. In general, large studies are needed to gain sufficient power to detect C.[25] Our study has by far the largest sample size; therefore it is possible that previous studies did not detect an influence of C due to the smaller sample sizes. However, these studies report very low, sometimes even negative DZ correlations and these cannot be explained by small sample sizes.[26] Other explanations for the difference between our findings and the findings of previous studies on the importance of the shared environment may include different approaches: we used a liability threshold model, whereas other studies used a continuous variable.

No specific risk factors were identified that explained the shared environmental contribution in this cohort. No association between birth cohort, socioeconomic status, parental age, maternal tobacco and alcohol consumption during pregnancy, paternal smoking behavior during pregnancy and gestational age and CWC and EU item scores was observed. The only predictor that reached significance in our study was gestational age. The association between preterm birth

and difficult temperament has been described before, however, in our study, its effect on the A, C and E loadings was very small.[27] Most studies investigating specific environmental factors that are shared between family members, like parental style and family functioning, have not established an association with temperament measures.[28]

One possible explanation for large effects of C is 'rater bias'. This is the systematic error that occurs when raters consistently over- or underestimate behavioral scores. If a rater then has to report on more than one child, scores can become correlated due to characteristics of the rater and in an ACE model this will appear like 'C'.[29] To explore this form of rater bias, the influence of maternal personality characteristics on CWC and EU ratings was estimated. A significant association was found only for Neuroticism and EU. However, this effect did not explain our large estimate of the influence of C. The model including Neuroticism of the mother yielded slightly lower A loadings and not lower C loadings. It is therefore probable that the influence of Neuroticism of the mother on EU ratings is due to genes that are shared between the mother and her children. In addition, the bivariate analysis showed that the correlation of CWC and EU (r=0.36), was for 50% and 48% due to shared genetic factors and for 42% and 44% due to shared environment influencing both traits, in boys and girls respectively. If a large proportion of our estimate of shared environmental influence would be due to rater bias one would expect a higher correlation between the items, and the covariance between the items to be mostly explained by the shared environment that is common to both traits. Still, the large correlated influences of C on both items may reflect rater characteristics that we did not account for in this study.

CWC and EU at age two were found to be predictive of internalizing, externalizing and attention problems at age 7. Several previous studies have investigated the association between difficult temperament or emotionality and later behavioral problems. Significant associations have been reported in many studies, but the correlations reported differ widely.[3-7] This might be due to the differences across studies with regard to age of measurement and instruments used. A more comprehensive comparison might be found in the literature on crying behavior in the first year of life conceptualized as 'regulatory problems'. A recent meta-analysis on the association of regulatory problems in the first year of life and later behavioral problems reported effect sizes of .50 (95% CI 0.27 to 0.73), .56 (95% CI 0.31 to 0.82), and.42 (95% CI 0.06 to 0.77) when comparing children with and without crying problems on ratings of internalizing, externalizing

and attention problems later in life.[30] Altogether, the effect sizes found in the present study seem slightly smaller but comparable to the effect sizes found in previous studies.

A limitation of the present study lies in the measurement instrument used, as only two items were assessed. However, the vast size of the cohort under investigation and its embedding in a longitudinal study design is a unique advantage of the present study, as it allowed us to test for factors possibly underlying the shared environment of twins. Furthermore, the longitudinal aspect of the study made it possible to establish that CWC and EU are predictive of later behavioral problems and thus likely capture an important aspect of infant behavior.

Much debate has been spent on the issue of the validity of parental ratings versus observer or laboratory ratings of child behavior. Although these ratings tend to differ substantially, parental reports have the overwhelming advantage of in depth knowledge of the child's behavior over different time points and situations and are therefore still considered of utter importance. Moreover, parent rated data can be collected in large numbers, whereas laboratory measures are always limited in size due to practical infeasibilities.

A somewhat surprising result emerged from tests of prevalence and thresholds: females from DZ opposite-sex pairs were rated as less easily upset and less often crying without a cause than female same-sex twins. This effect was not seen in their brothers. One could consider the possibility that having a male co-twin is protective for girls, but the observed effect might also represent a rater effect. Rietveld et al. reported a comparable phenomenon with regard to hyperactivity and attention problems; opposite-sex female twins were rated less hyperactive than same-sex female twins.[31] As these behaviors are more common in boys than in girls, this may cause parents to rate the female co-twin as less difficult when there is a brother as a comparison, a so called contrast effect. However, overall sex differences with regard to negative emotionality were not found in the present study nor in a large meta-analysis that assessed sex differences in temperament, including Negative Affectivity.[32] Despite the absence of sex differences, a contrast effect seems the most likely explanation for our findings.

References

1. Reijneveld SA, Brugman E, Hirasing RA. Excessive infant crying: the impact of varying definitions. Pediatrics. 2001;108:893-897.

2. Forsyth BW, Leventhal JM, McCarthy PL. Mothers' perceptions of problems of feeding and crying behaviors. A prospective study. Am J Dis Child. 1985;139:269-272.

3. Bates JE, Olson SL, Pettit GS, Bayles K. Dimensions of individuality in the mother-infant relationship at six months of age. Child Dev. 1982;53:446-461.

4. Caspi A, Silva PA. Temperamental qualities at age three predict personality traits in young adulthood: longitudinal evidence from a birth cohort. Child Dev. 1995;66:486-498.

5. Gjone H, Stevenson J. A longitudinal twin study of temperament and behavior problems: common genetic or environmental influences? J Am Acad Child Adolesc Psychiatry. 1997;36:1448-1456.

6. Guerin DW, Gottfried AW, Thomas CW. Difficult temperament and behaviour problems: a longitudinal study from 1.5 to 12 years. Int J Behav Dev. 1997;21:71-90.

7. Mun EY, Fitzgerald HE, Von Eye A, Puttler LI, Zucker RA. Temperamental characteristics as predictors of externalizing and internalizing child behavior problems in the contexts of high and low parental psychopathology. Infant Ment Health J. 2001;22:393-415.

8. Saudino KJ. The development of temperament from a behavioral genetics perspective. Adv Child Dev Behav. 2009;37:201-231.

9. Goldsmith HH, Buss KA, Lemery KS. Toddler and childhood temperament: expanded content, stronger genetic evidence, new evidence for the importance of environment. Dev Psychol. 1997;33:891-905.

10. Goldsmith HH, Lemery KS, Buss KA, Campos JJ. Genetic analyses of focal aspects of infant temperament. Dev Psychol. 1999;35:972-985.

11. Saudino KJ, McGuire S, Reiss D, Hetherington EM, Plomin R. Parent ratings of EAS temperaments in twins, full siblings, half siblings, and step siblings. J Pers Soc Psychol. 1995;68:723-733.

12. Heath AC, Kendler KS, Eaves LJ, Martin NG. Evidence for genetic influences on sleep disturbance and sleep pattern in twins. Sleep. 1990;13:318-335.

13. Jang KL, Livesley WJ, Vernon PA. Heritability of the big five personality dimensions and their facets: A twin study. J Pers. 1996;64:577-591.

14. Neale MC, Lubke G, Aggen SH, Dolan CV. Problems with using sum scores for estimating variance components: Contamination and measurement noninvariance. Twin Res Hum Genet. 2005;8:553-568.

15. Bartels M, van Beijsterveldt CE, Derks EM, et al. Young Netherlands Twin Register (Y-NTR): a longitudinal multiple informant study of problem behavior. Twin Res Hum Genet. 2007;10:3-11.

16. Rietveld MJ, van Der Valk JC, Bongers IL, Stroet TM, Slagboom PE, Boomsma DI. Zygosity diagnosis in young twins by parental report. Twin Res. 2000;3:134-141.

17. CBS. Standard classification of occupations. Heerlen/Voorburg: Central Bureau of Statistics; 2001

18. Erikson R, Goldthorpe J, Portocarero L, . Intergenerational class mobility in three Western European societies: England, France and Sweden. Br J Sociol 1979;30:415–441.

19. Costa PT, McCrae RR. Professional manual: Revised Neo Personality Inventory (Neo-PI-R) and NEO Five-Factor-Inventory (Neo-FFI). Odessa, FL: Psychological Assessment Resources; 1992.

20. Achenbach TM. Manual for the Child Behavior Checklist/4-18 and 1991 Profile. Burlington, VT: University of Vermont, Department of Psychiatry; 1991.

21. Verhulst FC, Van der Ende, J., Koot, H. M. Handleiding voor de CBCL/4-18. Rotterdam: Afdeling Kinder- en Jeugdpsychiatrie, Sophia Kinderziekenhuis / Academisch Ziekenhuis Rotterdam / Erasmus Universiteit Rotterdam; 1996.

22. Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. Nat Rev Genet. 2002;3:872-882.

23. Martin N, Boomsma D, Machin G. A twin-pronged attack on complex traits. Nat Genet. 1997;17:387-392.

24. Cohen J. Statistical power analyses for the behavioral sciences. 2nd ed. Hillsdale, NY: Erlbaum; 1988.

25. Posthuma D, Boomsma DI. A note on the statistical power in extended twin designs. Behav Genet. 2000;30:147-158.

26. Saudino KJ, Cherny SS, Plomin R. Parent ratings of temperament in twins: explaining the 'too low' DZ correlations. Twin Res. 2000;3:224-233.

27. Washington J, Minde K, Goldberg S. Temperament in preterm infants: style and stability. J Am Acad Child Psychiatry. 1986;25:493-502.

28. Daniels D, Plomin R, Greenhalgh J. Correlates of difficult temperament in infancy. Child Dev. 1984;55:1184-1194.

29. Bartels M, Boomsma DI, Hudziak JJ, van Beijsterveldt TC, van den Oord EJ. Twins and the study of rater (dis)agreement. Psychol Methods. 2007;12:451-466.

30. Hemmi MH, Wolke D, Schneider S. Associations between problems with crying, sleeping and/or feeding in infancy and long-term behavioural outcomes in childhood: a meta-analysis. Arch Dis Child. 2011;96:622-629.

31. Rietveld MJ, Hudziak JJ, Bartels M, van Beijsterveldt CE, Boomsma DI. Heritability of attention problems in children: I. cross-sectional results from a study of twins, age 3-12 years. Am J Med Genet B Neuropsychiatr Genet. 2003;117B:102-113.

32. Else-Quest NM, Hyde JS, Goldsmith HH, Van Hulle CA. Gender differences in temperament: a meta-analysis. Psychol Bull. 2006;132:33-72.

Chapter 3

Evidence for a causal association of low birth weight and attention problems

Groen-Blokhuis MM, Middeldorp CM, van Beijsterveldt CEM, Boomsma DI. Evidence for a causal association of low birth weight and attention problems. J Am Acad Child Adolesc Psychiatry 2011;50:1247-1254.

Abstract

Objective. Low birth weight (LBW) is associated with attention problems (AP) and attention deficit hyperactivity disorder (ADHD). The etiology of this association is unclear. We investigate if there is a causal influence of birth weight (BW) on AP and whether the BW effect is mediated by catch-up growth (CUG) in low BW children.

Method. Longitudinal data from > 29,000 twins registered with the Netherlands Twin Register with BW ≥1,500 grams and gestational age (GA) ≥ 32 weeks were analyzed with the co-twin control method. Hyperactivity and attention problems were assessed at ages 3, 7, 10, and 12 years; weight was assessed at birth and age 2.

Results. Children in the lowest BW category of 1,500-2,000 grams scored 0.18–0.37 standard deviations (SD) higher on AP than children in the reference category of 3,000–3,500 grams. This effect was present in term-born and preterm born children. Importantly, in BW discordant monozygotic (MZ), dizygotic (DZ), and unrelated (UR) pairs, the child with the lower BW scored higher on hyperactivity and attention problems than the child with the higher BW and within-pair differences were similar for MZ, DZ, and UR pairs. This pattern is consistent with a causal effect of BW on AP. MZ and DZ twin pairs concordant for LBW but discordant for CUG showed similar AP scores thus ruling out any effect of CUG on AP.

Conclusions. These results strongly indicate that the association of birth weight and AP represents a causal relationship. The effects of BW are not explained by CUG in LBW children.

Introduction

With a prevalence of 3%–10%, attention deficit and hyperactivity disorder (ADHD) forms a major burden for society and causes considerable impairment in the lives of children and adults.[1, 2] Twin studies have estimated the heritability of ADHD in children to be at least 60%.[1, 3] This implies that up to 40% of the variance of ADHD can be explained by other factors. Specific risk factors that have consistently been associated with attention problems (AP) and ADHD are male sex, perinatal trauma, maternal smoking during pregnancy, and low birth weight (LBW) (< 2500 grams).[4]

This paper focuses on the relationship between birth weight (BW) and AP. A recent meta-analysis reported children who were born very preterm (gestational age (GA) < 32 weeks) and/or with a very low birth weight (< 1500 grams, VLBW) to score respectively 0.43 and 0.59 SD higher on parent and teacher ratings of AP compared to controls.[5] Less attention has been devoted to the influence of less extreme forms of prematurity and LBW on symptoms of AP. However, studies that included the complete BW distribution in their analysis found the effect to extend over all BW categories below the average.[6, 7]

The question arises how the association between BW and AP originates. Does LBW increase the risk for AP? Or do other factors, such as GA, socioeconomic status, maternal stress, an unfavorable lifestyle during pregnancy or genetic factors, increase the risk for both LBW and AP? Twin studies provide a unique opportunity to investigate which mechanisms underlie an observed association. Monozygotic (MZ) twins share almost all their genes, while dizygotic (DZ) twins share on average 50% of their segregating genes. Both types of twins share many environmental factors, including SES, GA, and exposure to smoking during pregnancy. The co-twin control design provides the possibility to test for an association while controlling for genetic and environmental factors.[8, 9]

Two twin studies have previously investigated the association between BW and AP. A study in 1,480 twin pairs from the Swedish Twin Registry and a study in 2,007 MZ twin pairs from the Twin Early Development Study showed that the association between BW and AP was not due to common genetic or environmental factors but rather to a causal effect of BW on AP.[10, 11] For the

present study a much larger dataset was available consisting of 14,789 twin pairs who have been followed longitudinally from birth on.

Uniquely in this sample, it was possible to address the question whether the effect of LBW on AP is mediated by the effects of later catch-up growth (CUG). CUG describes the gain in weight SD score over time in the fist years after birth. LBW children frequently receive special nutritional programs in order to reach CUG. Although CUG has many positive effects, recent studies have reported negative effects of rapid CUG on diabetes, hypertension and, more recently, IQ.[12-14] We hypothesize that the BW effect on AP might be explained by CUG. To our knowledge, this is the first study to test the effect of CUG on AP.

The aim of the current study was to investigate the association of BW and AP along the entire distribution of BWs, to apply the co-twin control method to examine the mechanism underlying the BW-AP relationship and to addresses the question whether CUG in the first two years of life is causally related to later AP and possibly explains the BW-AP association.

Method

Subjects. Children included in this study are registered with the Netherlands Twin Register (NTR), established at the VU University Amsterdam in 1987. At birth, parents of multiples are invited to participate in longitudinal survey studies. A first survey is sent out after registration and later surveys are collected when the twins are 2, 3, 5, 7, 10, and 12 years old. For a more detailed description of the cohort and the data collection see Bartels et al.[15]

The twins included in this study were born between 1986 and 2003. Data on BW and AP were available for 16,398 twin pairs. Twin pairs were excluded if one of the children suffered from a severe handicap that interfered with daily functioning (n= 415 pairs) or if there were no data on GA (n = 74 pairs). As the focus of the study was on the less extreme values of BW and GA, 1,120 twin pairs with a GA less than 32 weeks or a BW smaller than 1,500 grams were excluded from the main analysis.[16] The final sample thus consisted of 14,789 twin pairs including 13,371 pairs with data available at age 3, 8,084 at age 7, 5,367 at age 10, and 4,578 at age 12. The number of twin pairs with available data decreases with increasing age due to non-response and the fact that not all children have reached the particular ages under study. Response rates of the questionnaires were 84% at age 1, 70% at age 3, 58% at age 7, 56% at age 10, and 50% at age 12.

Zygosity. For 14% of same-sex twins, zygosity was based on the results of DNA or blood group typing.[17,18] For the remaining pairs, zygosity was determined by a set of questions on twin similarity, that was included in longitudinal surveys. Of the 14,789 twin pairs included, there were 2,269 monozygotic male (MZM) pairs, 2,600 dizygotic male (DZM) pairs, 2,562 monozygotic female (MZF) pairs, 2,363 dizygotic female (DZF) pairs and 4,995 dizygotic opposite-sex (DOS) pairs. Based on DNA assigned 'true' zygosity, we looked at the percentage of twins correctly classified by questionnaire items as a function of their true zygosity and their concordant/discordant status. Across surveys, the percentage of correctly classified twin pairs was 97.3% for MZ concordant twins and 93.7% for DZ concordant twins. For MZ discordant pairs, 96.2% was correctly classified and 94.5% of DZ discordant pairs was correctly classified. These differences were not significant and discordance status clearly is not associated with misclassification.

Birth weight and gestational age. In the first survey that is sent out after registration, mothers are asked to fill out the BWs of their twins as assessed in the hospital and the duration of the twin pregnancy. Gestational age was rounded at half weeks. Twin pairs were classified as BW discordant if the BW of the smallest twin was at least 15% lower than the BW of the heaviest twin or if there was a BW difference of at least 400 grams.[11, 19] To exclude the most extreme cases who may be suffering from transfusion syndrome, 300 twin pairs with a BW difference of more than 40% or more than 1000 grams were excluded from the analysis of BW discordant pairs.

Catch-up growth. In the surveys at age 2 and 3 mothers are asked to fill out the weights of their twins as assessed by the Dutch National Health Services at regular intervals. For weight at age 2 we selected the measurement between 18 and 30 months that was closest to 24 months. Twin pairs were classified as concordant for LBW if both twins had a BW < 2500 grams and if the BW of the smallest twin was less than 10% lower than the BW of the heaviest twin. Discordance for CUG was defined as a gain in weight SD score over the first 2 years of > 0.67 SD in one twin and ≤ 0.50 SD in the other twin.[20] Weight data at age 2 were standardized with the sex and age specific Dutch growth charts for the general population from 1997 using the software package Growth analyser 3.[21, 22] Weight at birth was standardized using the reference data of Niklasson et al. with correction for gestational age.[23]

Attention Problems. An age appropriate Dutch version of the Achenbach System of Empirically Based Assessment (ASEBA) is included in the NTR surveys at age 3, 7, 10, and 12.[24, 25] The Overactive scale at age 3 contains 5 items and the Attention Problems scale at age 7, 10, and 12 contains 11 items describing both hyperactive and inattentive behaviors. As the number of items differs over the CBCL scales, T-scores were calculated to facilitate interpretation. A T-score of 50 represents the mean and a deviation of 10 points from the mean represents a deviation of 1 SD. A T-score above 65 indicates clinically significant problem behavior. T-scores were calculated separately for boys and girls. We report on maternal AP ratings; results were similar when analyzing ratings from fathers (see Table S1).

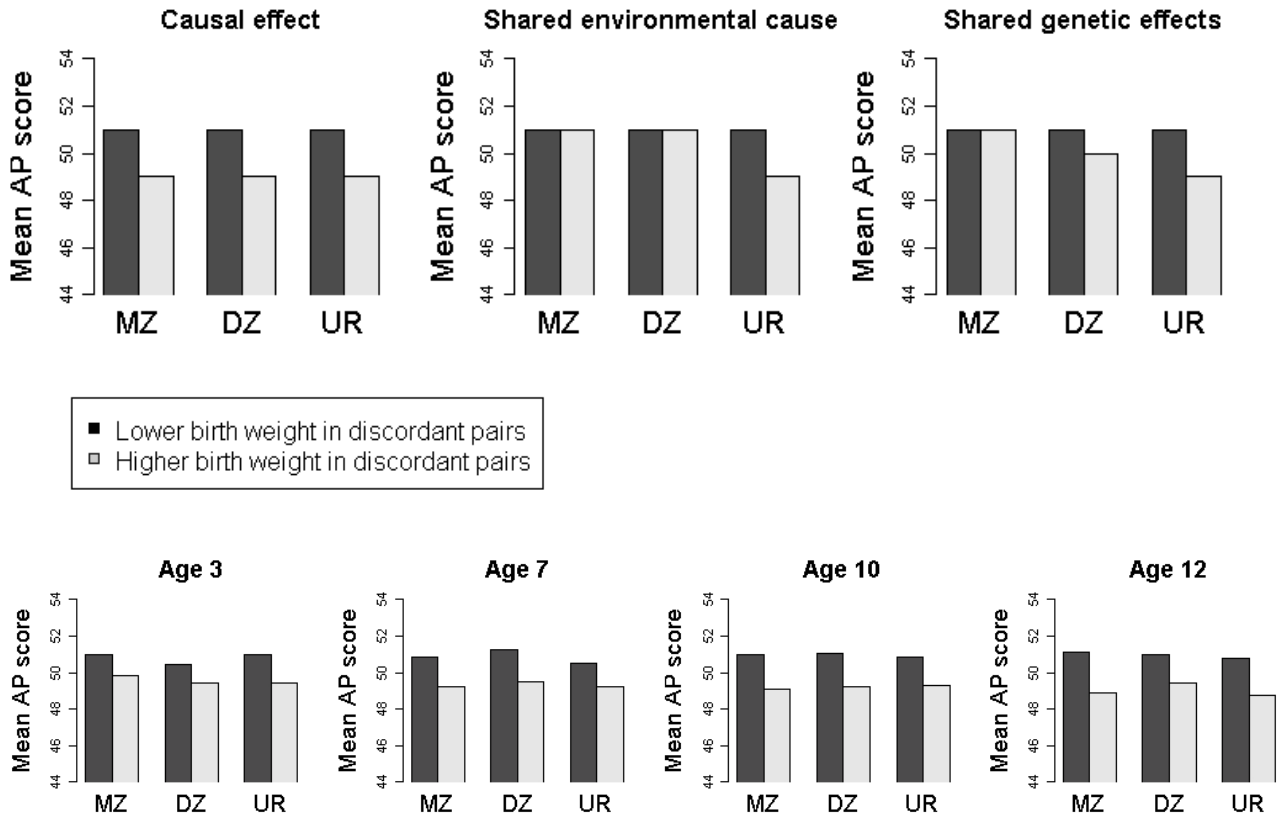
BW&AP. Mean T-scores on the AP scale at age 3, 7, 10, and 12 were obtained for each BW category. Data from two individuals within a twin pair are not independent. Therefore, mean AP scores for all BW categories are given for the following 4 groups: first born males, second born

males, first born females, and second born females. A linear regression was performed to test whether BW significantly predicted AP scores. Effect sizes were calculated in terms of Cohen's *d:* the difference between 2 means divided by the pooled SD for those means.[26]

Co-twin control method. The co-twin control method addresses issues of causality.[8, 9] In this design, the association between BW and AP is tested in three groups: MZ twin pairs, DZ twin pairs, and a sample of UR pairs discordant for BW. Monozygotic (MZ) twins share almost all their DNA while dizygotic (DZ) twins share on average 50% of their segregating genes. Both types of twins share many environmental factors including SES, GA, and exposure to smoking during pregnancy, while the UR pairs share genetic and environmental factors at a random level. The upper figure in Figure 1 depicts an example of the expected mean AP scores in BW discordant pairs from the MZ, DZ, and UR group under the three possible models that explain the association between BW and AP. In the case of a causal model, all within-pair differences in BW will result in within-pair differences in AP. Hence, the individuals with the higher BW in a pair will show similarly lower AP scores in the MZ, DZ, and UR pairs. If the association between BW and AP is due to an environmental factor that is shared between twins, only UR pairs will differ in their AP scores with the individuals with the higher BW showing lower AP scores than the individuals with the lower BW. Due to the fact that twins have an equal exposure to the environmental factor causing both LBW and AP, the twins with the higher BW in a pair will show the same mean AP scores as the co-twins with the lower BW. Finally, if the association is due to shared genetic factors, the difference in BW will again be present in the UR pairs. However, MZ twins discordant for BW will share the causal genetic factors for AP and hence, the MZ twins with the higher BW in a pair are expected to show the same mean AP score as the co-twins with the lower birth weight, whereas the DZ group will show an intermediate pattern.

Figure 1. Expected and observed patterns of attention problems (AP) scores in monozygotic (MZ) and dizygotic (DZ) twin pairs and unrelated (UR) pairs discordant for birth weight (BW). Upper figure: Expected patterns under 3 theoretical models: 1) Association between BW and AP due to a causal effect of BW on AP 2) Association between BW and AP due to an environmental factor that influences both BW and AP and is shared between twins 3) Association between BW and AP due to genetic factors that influence both BW and AP. Lower figure: Observed mean T-

scores at the Attention Problems (AP) scale in monozygotic (MZ) and dizygotic (DZ) twin pairs and unrelated (UR) pairs discordant for BW at age 3, 7, 10, and 12.



As sex is an important determinant of both BW and AP, only same-sex pairs were included in the DZ BW discordant group. A group of UR pairs discordant for BW was created as follows. Pairs of unrelated individuals were drawn from the group of twins that were not discordant for BW and matched to the BW discordant MZ and DZ pairs based on sex, GA (rounded at 1 week) and BW category (categories of 100 grams were used). Thus, the group of UR pairs has the same characteristics with regard to BW discordance and GA as the MZ and DZ BW discordant groups.

To test for mean differences within BW discordant twin pairs, a paired t-test was performed in MZ and DZ twin pairs and an unpaired t-test was performed in UR pairs. To test for the different explanatory models, a linear regression was performed in which it was tested whether MZ, DZ and UR group membership (coded as 0, -1 and -2) significantly predicted AP scores in the group of individuals with the higher birth weight in a pair and in the group of individuals with the

lower birth weight in a pair. As an additional test, it was tested whether group membership (MZ, DZ or UR) modified the effect of within-pair BW differences on AP, by including the interaction term in the linear regression.

Catch-up growth. LBW and CUG are highly correlated. To disentangle the effects of LBW and CUG on AP, 190 same-sex twin pairs concordant for LBW yet discordant for CUG were selected. To test for the effect of CUG on AP, mean differences in AP scores at age 3, 7, 10, and 12 between the twin with the higher CUG and the co-twin with the lower CUG were tested with a paired t-test.

Results

BW&AP. The mean AP scores at age 3, 7, 10, and 12 for each BW category are shown in Table 1. As the results were roughly the same for first and second born twins, only the results for first born males and females are shown. Clearly, AP scores decrease with increasing BW up to BWs of 3,500 grams in all age groups. The association between BW and AP was tested in a linear regression; BW significantly predicted AP scores at all ages ($p < 0.05$). Taking the BW category of 3,000–3,500 grams as the reference category, children with a BW of 1,500–2,000 grams scored 0.18–0.37 SD higher on the AP scale. In line with this, the percentage of children scoring above the clinical cutoff decreases with increasing BW (Table 1). In a subsample of children who were term-born (GA ≥ 37 weeks) the same pattern was found with somewhat larger differences of 0.27–0.70 SD (see Table S2).

Table 1. Mean and SD of T-scores of maternal ratings of the Attention problems (AP) scale and the *percentage of children with a T score > 65* at age 3, 7, 10, and 12 for each birth weight (BW) category in first born males (M) and females (F) with gestational age ≥32 weeks.

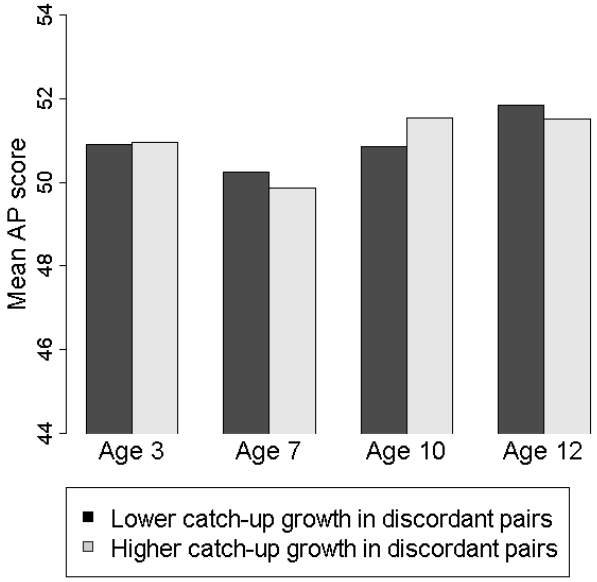
Sex	BW	n	AP 3	n	AP 7	n	AP 10	n	AP 12
M	1,500–1,999	598	51.3 (10.2) *8.7%*	348	51.8 (10.9) *14.7%*	209	50.6 (10.3) *9.6%*	173	51.4 (10.0) *9.8%*
	2,000–2,499	1,817	50.6 (9.9) *7.4%*	1,126	50.6 (10.1) *9.1%*	723	50.6 (10.3) *7.9%*	629	51.0 (11.0) *11.4%*
	2,500–2,999	2,627	50.0 (9.8) *6.6%*	1,540	49.8 (9.7) *9.0%*	1,029	50.1 (9.9) *6.5%*	876	49.8 (9.7) *8.1%*
	3,000–3,499	1,384	49.4 (9.7) *6.4%*	870	49.0 (9.5) *8.3%*	562	48.4 (9.5) *5.9%*	497	49.0 (10.0) *7.6%*
	≥3,500	235	48.4 (9.7) *5.1%*	131	49.1 (9.7) *6.9%*	87	50.2 (9.6) *4.1%*	74	49.4 (8.2) *4.0%*
F	1,500–1,999	740	51.5 (9.9) *11.4%*	457	51.9 (10.5) *9.6%*	308	51.7 (10·9) *9.4%*	265	51.5 (9.6) *12.1%*
	2,000–2,499	2,131	50.7 (10.1) *11.4%*	1,271	51.2 (10.8) *9.9%*	885	50.8 (10.7) *9.0%*	731	51.0 (10.7) *12.9%*
	2,500–2,999	2,637	49.6 (9.9) *9.0%*	1,646	49.6 (9.8) *7.7%*	1,099	49.7 (9.6) *6.4%*	939	49.4 (9.7) *8.4%*

	3,000–3,499	990	48.9 (9.7) *8.6%*	605	49.3 (9.8) *7.8%*	398	49.0 (9.3) *6.0%*	347	49.1 (9.5) *7.8%*
	≥3,500	133	48.6 (9.0) *6.0%*	78	49.6 (11.3) *12.8%*	56	47.5 (9.0) *3.6%*	44	46.1 (8.3) *6.8%*

Attention problems in birth weight discordant pairs. 1,258 MZ twin pairs were discordant for BW. AP data were available for 1,133 of these twin pairs at age 3, 723 at age 7, 500 at age 10, and 441 at age 12. There were 1,587 same-sex DZ twin pairs discordant for BW. AP data were available for 1,425 of these twin pairs at age 3, 846 at age 7, 528 at age 10, and 444 at age 12. The average BW differences were similar in the MZ, DZ and UR pairs, with mean BW differences of respectively 21%, 20% and 20% and mean absolute differences of 562, 581 and 560 grams. The average AP scores in the BW discordant MZ, DZ and UR pairs are depicted in the lower figure of Figure 1. At all ages, the average AP score of the individuals with the lower BW in a pair was higher than the average score of the individuals with the higher BW in a pair, with differences between 0.10–0.24 SD. These differences were significant in all pairs at all ages of measurement ($p < 0{\cdot}01$). A linear regression was performed to test for the different explanatory models. At all ages of measurement, group membership (MZ, DZ or UR) did not predict AP scores in both the higher birth weight individuals and the lower birth weight individuals in a pair. In line with this, group membership (MZ, DZ or UR) did not significantly interact with the effect of within-pair BW differences on AP scores at all ages ($p > 0{\cdot}05$).

Catch-up growth & attention problems. 52 MZ and 138 same-sex DZ twin pairs were concordant for LBW and discordant for CUG. The AP scores of the twins with the higher CUG in a pair did not differ from the AP scores of the twins with the lower CUG in a pair, at all ages of measurement ($p > 0{\cdot}05$) (Figure 2).

Figure 2. Mean T-scores at the Attention Problems (AP) scale at age 3, 7, 10, and 12 in twin pairs concordant for low birth weight yet discordant for catch-up growth.



Discussion

In this large twin study (n = 14,789 twin pairs) there was a clear negative relationship between BW and AP. Children with a BW of 1,500–2,000 grams scored 0.18–0.37 SD higher on the AP scale than children in the reference category of 3,000–3,500 grams. This effect was present in term-born (GA ≥37 weeks) and preterm born children (GA 32–37 weeks). Within MZ, DZ, and UR BW discordant pairs, individuals with the lower BW in a pair scored 0.10–0.24 SD higher on the AP scale than the individuals with the higher BW in a pair. The effect of BW on AP was similar over the MZ, DZ, and UR groups, in line with a causal effect of BW on AP. We were able to identify a group of twin pairs who were concordant for LBW yet discordant for CUG. In this group no differences in longitudinally assessed AP scores was found.

The strength of the present study lies in the prospective data collection, the large sample size and the possibility to control for prematurity, genetic factors, maternal factors and catch-up growth within the co-twin control design.

Many studies have focused on the effects of VLBW and extreme prematurity on AP. In line with previous studies, the influence of BW on AP was present over the full range of BWs in our cohort.[6, 7] The clinical importance of this finding is considerable, as the number of children with moderately LBW (1,500–2,500 grams) is much larger than the number of children with VLBW.

The effect sizes observed in our study are slightly smaller than the effect sizes described previously.[5] These differences can be explained by the fact that children with BW < 1,500 grams were excluded. When all children with available BW and AP scores were included, differences in AP scores of 0.21–0.53 SD were found between children with a BW < 1,500 grams and children in the reference category of 3,000–3,500 grams.

Previous reviews have criticized the existent studies on the influence of LBW on AP since many of these studies did not control for the effect of prematurity, socioeconomic status, maternal tobacco and alcohol consumption, and genetic factors.[4] The fact that we found the effect of BW on AP in preterm and term-born children and, more importantly, that the effect was the same in MZ, DZ, and UR BW discordant pairs, makes clear that BW is a risk factor for AP independent from the above listed factors.

This raises the question how LBW might cause AP, i.e. is this a direct effect or is LBW a proxy for another risk factor causing AP? Animal studies and human postmortem studies suggest that growth restriction in utero, as reflected in LBW, contributes to later AP. These studies consistently showed an influence of intrauterine growth restriction (IUGR) on brain volume and development (reviewed by de Bie et al. and Schlotz et al.[27, 28]) Some of the animal studies are of particular importance for AP as they found an effect of IUGR on the striatum, a brain structure that is thought to play a key role in the pathogenesis of ADHD.[29, 30] Interestingly, one MRI study in MZ twins discordant for ADHD found smaller caudate volumes in affected twins, pointing to a deficit in frontal-striatal processing.[31] In another MRI study in MZ twins discordant for AP, the high-scoring twins showed a volume reduction of the inferior dorsolateral prefrontal cortex and decreased activation in the left and right temporal lobe areas during the color-word Stroop task compared to the low-scoring twins.[32, 33] The high-scoring twins were previously found to have a lower average BW than the AP low-scoring twins.[34] Together, these findings support the idea that BW differences in MZ twins reflect differential nourishment in utero, leading to impaired neurodevelopment. This is in line with the Developmental Origins of Health and Disease model (Barker's hypothesis), an influential hypothesis that states that an adverse environment during critical periods of fetal life increases the risk of a wide range of diseases later in life, including diabetes, cardiovascular disease, cancer, and neuropsychiatric disorders.[28, 35] A recent study[36] in newborn MZ twins obtained evidence for the hypothesis that epigenetic factors accumulated in utero can contribute to low birth weight and predisposition to complex diseases later in life. The number of twins in this study were small and more studies are needed to confirm that MZ twins can be discordant at birth for epigenetics and gene expression.[37]

It has been suggested that the number of children diagnosed with ADHD has increased over time, as expressed by a marked increase in medication use.[38] Interestingly, a recent study in more than 36 million term born children in the United States showed a decrease in BW over the period 1990–2005.[39] Based on our findings, an increase in ADHD might partly be explained by the decrease in BW.

Generalization from our study relies on the assumption that the liability to develop AP is influenced by similar factors in twins and singletons. Prematurity, LBW and IUGR occur more frequently in twins and, more importantly, the causes and effects of IUGR in twins might differ

from those in singletons. However, one of the reasons to assume that the latter could be the case is the fact that studies linking LBW to various outcomes in twins have sometimes lacked to find an association that was present at the population level. As we do find a clear association between BW and AP in our large cohort and in the MZ and DZ BW discordant twin pairs, this argument does not hold for the relationship between BW and AP. Important to notice in this regard is that twins do not differ from singletons with regard to behavioral problems, and specifically AP.[40,41] This suggests that an unfavourable uterine environment leading to LBW is as much a risk factor for AP in the twin population as it is in singletons.

Another limitation of our study is that inattention and hyperactive behavior were assessed by means of a structured questionnaire instead of a clinical diagnosis of ADHD. Although the latter directly relates to clinical practice, a liability approach in which AP is considered a continuous trait instead of a dichotomy has been shown to adequately represent AP on the population level and provides better statistical power. More importantly, the CBCL-AP scale converges with the results of clinical interviews covering the DSM-IV criteria.[42]

As all data were collected by questionnaire surveys in a population based cohort, non-random participation and drop-out over time is a concern. A study by Gielen et al. compared the gestational ages and birth weights in the NTR with a reference data set of all Dutch live born twins (NPR) and found that gestational age and birth weights of the NTR were higher then those of the NPR, although the differences were small: NPR: 35.9 (3.0) weeks and 2459 (615) grams; NTR: 36.5 (2.4) weeks and 2498 (550) grams.[16] Similarly, another study from the NTR found slightly higher overactive behavior scores at age 3 in children that did not participate in the surveys at age 7, 10 or 12.[43] These forms of selection bias might have led to an underestimation of the true effect sizes in the current study.

A final limitation is that the number of twin pairs concordant for LBW but discordant for CUG was still limited although the total sample size was large,.

To conclude, this study provides evidence for the influence of LBW on AP over the full range of BWs. The analysis in BW discordant twins shows that this effect is not accounted for by prematurity, SES, or tobacco and alcohol consumption during pregnancy. We hypothesize that deficient nourishment in utero leads to impaired neurodevelopment and the occurrence of AP.

Whereas recent studies have reported negative outcomes of rapid CUG, our analysis of the effects of CUG on AP indicate that catch-up growth is neither beneficial nor harmful with regard to AP. The decision to treat children with LBW with nutritional-enriched diets should be based on a careful consideration of the positive and negative effects of CUG; future research should aim to unravel the characteristics of healthy catch-up growth.

References

1. Faraone SV, Doyle AE. The nature and heritability of attention-deficit/hyperactivity disorder. Child Adolesc Psychiatr Clin N Am. 2001;10:299-316, viii-ix.
2. Rowland AS, Lesesne CA, Abramowitz AJ. The epidemiology of attention-deficit/hyperactivity disorder (ADHD): a public health view. Ment Retard Dev Disabil Res Rev. 2002;8:162-170.
3. Derks EM, Hudziak JJ, Boomsma DI. Genetics of ADHD, Hyperactivity and Attention Problems. In: Kim Y, ed. Handbook of Behavior Genetics. New York: Springer; 2009: 361-378.
4. Zappitelli M, Pinto T, Grizenko N. Pre-, peri-, and postnatal trauma in subjects with attention-deficit hyperactivity disorder. Can J Psychiatry. 2001;46:542-548.
5. Aarnoudse-Moens CS, Weisglas-Kuperus N, van Goudoever JB, Oosterlaan J. Meta-analysis of neurobehavioral outcomes in very preterm and/or very low birth weight children. Pediatrics. 2009;124:717-728.
6. Boulet SL, Schieve LA, Boyle CA. Birth Weight and Health and Developmental Outcomes in US Children, 1997-2005. Matern Child Health J. 2011;15:836-844.
7. Stein RE, Siegel MJ, Bauman LJ. Are children of moderately low birth weight at increased risk for poor health? A new look at an old question. Pediatrics. 2006;118:217-223.
8. Cederlof R. The Twin Method in Epidemiological Studies on Chronic Disease (doctoral dissertation). Stockholm, Sweden: University of Stockholm; 1966.
9. Middeldorp CM, Cath DC, Beem AL, Willemsen G, Boomsma DI. Life events, anxious depression and personality: a prospective and genetic study. Psychol Med. 2008;38:1557-1565.
10. Asbury K, Dunn JF, Plomin R. Birthweight-discordance and differences in early parenting relate to monozygotic twin differences in behaviour problems and academic achievement at age 7. Dev Sci. 2006;9:F22-F31.
11. Hultman CM, Torrang A, Tuvblad C, Cnattingius S, Larsson JO, Lichtenstein P. Birth weight and attention-deficit/hyperactivity symptoms in childhood and early adolescence: a prospective Swedish twin study. J Am Acad Child Adolesc Psychiatry. 2007;46:370-377.

12. Estourgie-van Burk GF, Bartels M, Hoekstra RA, Polderman TJ, Delemarre-van de Waal HA, Boomsma DI. A twin study of cognitive costs of low birth weight and catch-up growth. J Pediatr. 2009;154:29-32.
13. Huxley RR, Shiell AW, Law CM. The role of size at birth and postnatal catch-up growth in determining systolic blood pressure: a systematic review of the literature. J Hypertens. 2000;18:815-831.
14. Ong KK. Catch-up growth in small for gestational age babies: good or bad? Curr Opin Endocrinol Diabetes Obes. 2007;14:30-34.
15. Bartels M, van Beijsterveldt CE, Derks EM, et al. Young Netherlands Twin Register (Y-NTR): a longitudinal multiple informant study of problem behavior. Twin Res Hum Genet. 2007;10:3-11.
16. Gielen M, van Beijsterveldt CE, Derom C, et al. Secular trends in gestational age and birthweight in twins. Hum Reprod. 2010;25:2346-2353.
17. van Dijk BA, Boomsma DI, de Man AJ. Blood group chimerism in human multiple births is not rare. Am J Med Genet. 1996;61:264-268.
18. Rietveld MJ, van Der Valk JC, Bongers IL, Stroet TM, Slagboom PE, Boomsma DI. Zygosity diagnosis in young twins by parental report. Twin Res. 2000;3:134-141.
19. Patterson RM, Wood RC. What is twin birthweight discordance? Am J Perinatol. 1990;7:217-219.
20. Ong KK, Ahmed ML, Emmett PM, Preece MA, Dunger DB. Association between postnatal catch-up growth and obesity in childhood: prospective cohort study. BMJ. 2000;320:967-971.
21. Growth analyser 3 [computer program]. Rotterdam, the Netherlands: Dutch Growth Foundation; 2004.
22. Fredriks AM, van Buuren S, Burgmeijer RJ, et al. Continuing positive secular growth change in The Netherlands 1955-1997. Pediatr Res. 2000;47:316-323.
23. Niklasson A, Ericson A, Fryer JG, Karlberg J, Lawrence C, Karlberg P. An update of the Swedish reference standards for weight, length and head circumference at birth for given gestational age (1977-1981). Acta Paediatr Scand. 1991;80:756-762.
24. Achenbach TM. Manual for the Child Behavior Checklist/4-18 and 1991 Profile. Burlington, VT: University of Vermont, Dep of Psychiatry; 1991.

25. Verhulst FC, van der Ende J, Koot HM. Handleiding voor de CBCL/4-18. Rotterdam: Afdeling Kinder- en jeugdpsychiatrie, Sophia Kinderziekenhuis/Academisch Ziekenhuis Rotterdam/Erasmus Universiteit Rotterdam; 1996.
26. Cohen J. Statistical power analyses for the behavioral sciences. 2nd ed. Hillsdale, NY: Erlbaum; 1988.
27. de Bie HM, Oostrom KJ, Delemarre-van de Waal HA. Brain development, intelligence and cognitive outcome in children born small for gestational age. Horm Res Paediatr. 2010;73:6-14.
28. Schlotz W, Phillips DI. Fetal origins of mental health: evidence and mechanisms. Brain Behav Immun. 2009;23:905-916.
29. Mallard EC, Waldvogel HJ, Williams CE, Faull RL, Gluckman PD. Repeated asphyxia causes loss of striatal projection neurons in the fetal sheep brain. Neuroscience. 1995;65:827-836.
30. Mallard EC, Williams CE, Johnston BM, Gunning MI, Davis S, Gluckman PD. Repeated episodes of umbilical cord occlusion in fetal sheep lead to preferential damage to the striatum and sensitize the heart to further insults. Pediatr Res. 1995;37:707-713.
31. Castellanos FX, Sharp WS, Gottesman RF, Greenstein DK, Giedd JN, Rapoport JL. Anatomic brain abnormalities in monozygotic twins discordant for attention deficit hyperactivity disorder. Am J Psychiatry. 2003;160:1693-1696.
32. van 't Ent D, Lehn H, Derks EM, et al. A structural MRI study in monozygotic twins concordant or discordant for attention/hyperactivity problems: evidence for genetic and environmental heterogeneity in the developing brain. Neuroimage. 2007;35:1004-1020.
33. van 't Ent D, van Beijsterveldt CE, Derks EM, et al. Neuroimaging of response interference in twins concordant or discordant for inattention and hyperactivity symptoms. Neuroscience. 2009;164:16-29.
34. Lehn H, Derks EM, Hudziak JJ, Heutink P, van Beijsterveldt TC, Boomsma DI. Attention problems and attention-deficit/hyperactivity disorder in discordant and concordant monozygotic twins: evidence of environmental mediators. J Am Acad Child Adolesc Psychiatry. 2007;46:83-91.
35. Barker DJ. The developmental origins of well-being. Philos Trans R Soc Lond B Biol Sci. 2004;359:1359-1366.

36. Gordon L, Joo JH, Andronikos R, Ollikainen M, Wallace EM, Umstad MP, Permezel M, Oshlack A, Morley R, Carlin JB, Saffery R, Smyth GK, Craig JM. Expression discordance of monozygotic twins at birth: effect of intrauterine environment and a possible mechanism for fetal programming. Epigenetics. 2011;6:579-92
37. Bell JT, Spector TD. A twin approach to unraveling epigenetics. Trends Genet. 2011;27:116-25.
38. Robison LM, Skaer TL, Sclar DA, Galin RS. Is attention deficit hyperactivity disorder increasing among girls in the US? Trends in diagnosis and the prescribing of stimulants. CNS Drugs. 2002;16:129-137.
39. Donahue SM, Kleiman KP, Gillman MW, Oken E. Trends in birth weight and gestational length among singleton term births in the United States. Obstet Gynecol. 2010;115:357-364.
40. Ehringer MA, Rhee SH, Young S, Corley R, Hewitt JK. Genetic and environmental contributions to common psychopathologies of childhood and adolescence: a study of twins and their siblings. J Abnorm Child Psychol. 2006;34:1-17.
41. Gjone H, Novik TS. Parental ratings of behaviour problems: a twin and general population comparison. J Child Psychol Psychiatry. 1995;36:1213-1224.
42. Biederman J, Faraone SV, Doyle A, et al. Convergence of the Child Behavior Checklist with structured interview-based psychiatric diagnoses of ADHD children with and without comorbidity. J Child Psychol Psychiatry. 1993;34:1241-1251.
43. Rietveld MJ, Hudziak JJ, Bartels M, van Beijsterveldt CE, Boomsma DI. Heritability of attention problems in children: longitudinal results from a study of twins, age 3 to 12. J Child Psychol Psychiatry. 2004;45:577-588.

Supplementary Materials

Table S1. Mean and SD of T-scores of paternal ratings at the Attention problems (AP) scale *and the percentage of children with a T score > 65* at age 3, 7, 10, and 12 for each birth weight (BW) category in first born males (M) and females (F) with gestational age ≥ 32 weeks.

Sex	BW	n	AP 3	N	AP 7	n	AP 10	n	AP 12
M	1,500–1,999	406	51.3 (10.2) *7.9%*	260	51.7 (10.7) *9.2%*	148	52.4 (12·0) *12.2%*	135	51.3 (10.0) *8.9%*
	2,000–2,499	1,179	50.4 (10.2) *7.8%*	825	51.0 (10.3) *8.8%*	513	50.1 (10.3) *9.6%*	454	50.9 (10.8) *9.0%*
	2,500–2,999	1,754	49.7 (9.8) *6.0%*	1,155	49.7 (10.0) *6.8%*	747	49.8 (9.9) *8.2%*	641	49.6 (9.7) *7.6%*
	3,000–3,499	923	49.8 (10.0) *6.2%*	664	49.4 (9.5) *6.0%*	425	49.3 (9.5) *5.2%*	362	49.5 (9.6) *8.3%*
	≥3,500	151	48.9 (9.1) *4.0%*	96	47.9 (8.7) *5.2%*	68	49.6 (8.1) *4.4%*	60	48.8 (8.2) *10.0%*
F	1,500–1,999	474	51.0 (10.0) *9.5%*	327	51.0 (10.2) *9.5%*	219	52.1 (12.8) *14.6%*	188	51.3 (10.4) *10.1%*
	2,000–2,499	1,395	50.6 (10.1) *9.8%*	925	51.2 (10.9) *11.1%*	628	50.8 (10.1) *10.5%*	515	50.6 (10.2) *8.9%*
	2,500–2,999	1,763	49.9 (9.9) *7.9%*	1,196	49.5 (9.6) *7.6%*	804	49.1 (9.4) *7.8%*	702	49.7 (10.0) *8.4%*
	3,000–3,499	652	49.0 (10.0) *7.7%*	464	49.1 (9.8) *5.6%*	286	49.3 (9.1) *7.0%*	257	49.3 (9.4) *6.6%*
	≥3,500	86	47.8 (8.6) *5.8%*	58	50.7 (14.5) *13.8%*	38	48.0 (8.8) *7.9%*	35	46.8 (8.9) *8.6%*

Table S2. Mean and SD of T-scores of maternal ratings at the Attention problems (AP) scale and *the percentage of children with a T score > 65* at age 3, 7, 10, and 12 for each birth weight (BW) category in first born males (M) and females (F) with gestational age ≥ 37 weeks.

Sex	BW	n	AP 3	n	AP 7	n	AP 10	n	AP 12
M	1,500–1,999	72	52.3 (10.7) *12.5%*	41	53.5 (10.6) *17.1%*	28	53.1 (8.7) *14.3%*	27	51.5 (7.6) *3.7%*
	2,000–2,499	715	50.8 (9.8) *7.6%*	458	50.8 (10.0) *9.2%*	284	51.4 (10.2) *8.1%*	242	52.6 (12.0) *14.5%*
	2,500–2,999	2,040	49.9 (9.7) *6.4%*	1,188	49.9 (9.8) *9.4%*	799	50.1 (9·9) *6.3%*	678	50.0 (9.7) *8.0%*
	3,000–3,499	1,278	49.5 (9.7) *6.5%*	809	49.0 (9.3) *7.9%*	525	48.4 (9.3) *5.5%*	466	48.9 (9.6) *7.3%*
	≥3,500	226	48.6 (9.9) *5.3%*	128	49.2 (9.8) *7.0%*	85	50.4 (9.7) *5.9%*	72	49.4 (8.2) *4.2%*
F	1,500–1,999	98	52.4 (9.0) *12.2%*	63	52.9 (12.0) *12.7%*	39	52.6 (9.1) *7.7%*	30	52.7 (10.5) *20.0%*
	2,000–2,499	1,001	50.6 (10.1) *11.1%*	597	51.1 (10.5) *8.9%*	446	51.1 (10.8) *9.9%*	351	51.5 (11.0) *12.8%*
	2,500–2,999	2,188	49.7 (10.0) *9.2%*	1,382	49.6 (9.8) *8·0%*	913	49.7 (9.7) *6.4%*	786	49.8 (9.9) *8.9%*
	3,000–3,499	945	48.8 (9.7) *8.6%*	584	49.0 (9.6) *7.2%*	382	48.9 (9.2) *5.5%*	335	49.0 (9.4) *7.5%*
	≥3,500	131	48.6 (9.4) *6.1%*	76	49.4 (11.0) *11.8%*	54	47.4 (9.0) *3.7%*	35	46.1 (8.3) *6.8%*

Chapter 4

A prospective study of the effects of breastfeeding and FADS2 polymorphisms on cognition and hyperactivity/attention problems

Groen-Blokhuis MM, Franić S, van Beijsterveldt CEM, et al. A prospective study of the effects of breastfeeding and FADS2 polymorphisms on cognition and hyperactivity/attention problems. Am J Med Genet B Neuropsychiatr Genet 2013;162B:457-465.

Abstract

Breastfeeding has been associated with improved cognitive functioning. There is a beneficial effect on IQ, and possibly on associated phenotypes such as attention problems. It has been suggested that the effect on IQ is moderated by polymorphisms in the FADS2 gene, which is involved in fatty acid metabolism. In this study we tested the relation between breastfeeding and FADS2 polymorphisms on the one hand and IQ, educational attainment, overactivity, and attention problems on the other hand. IQ at age 5, 7, 10, 12, and/or 18 (n=1,313), educational attainment at age 12 (n=1,857), overactive behavior at age 3 (n=2,560), and attention problems assessed at age 7, 10, and 12 years (n=2,479, n=2,423, n=2,226) were predicted by breastfeeding and two SNPs in FADS2 (rs174575 and rs1535). Analyses were performed using structural equation modelling. After correction for maternal education, a main effect of breastfeeding was found for educational attainment at age 12 and overactive behavior at age 3. For IQ, the effect of breastfeeding across age was marginally significant (p=0.05) and amounted to 1.6 points after correcting for maternal education. Neither a main effect of the FADS2 polymorphisms nor an interaction with breastfeeding was detected for any of the phenotypes. This developmentally informed study confirms that breastfeeding is associated with higher educational attainment at age 12, less overactive behavior at age 3 and a trend toward higher IQ after correction for maternal education. In general, the benefits of breastfeeding were small and did not interact with SNPs in FADS2.

Introduction

The positive effects of breastfeeding for the newborn have been shown in many studies and have led the WHO to promote breastfeeding worldwide.[1] A positive association between breastfeeding and cognition has been reported in a substantial number of studies, as reviewed by Horta et al.[1] However, studies examining the effect of breastfeeding on IQ are complicated by several confounding effects, most importantly maternal IQ, socioeconomic status (SES) and maternal education.[2, 3] Although a meta-analysis and a large randomized trial reported an effect of breastfeeding on cognition independent of several confounders, other studies, including a meta-analysis and a sibling pairs analysis, did not find an association after control for confounding effects.[2-5]

During childhood, IQ consistently shows a negative association with Attention Deficit/Hyperactivity Disorder (ADHD) and attention problems (AP).[6] The effect of breastfeeding on AP and ADHD is less well studied than the effect of breastfeeding on IQ. In a case-control study in 100 children aged 4-11, ADHD cases were breastfed for a significantly shorter period than controls.[7] Furthermore, a prospective cohort study in 500 children found that long-term breastfeeding was associated with fewer ADHD symptoms and improved executive functioning after correction for sociodemographic characteristics of the parents.[8] A study of 1,287 boys aged 6-13 found that inattention was significantly higher among bottle-fed boys, but this effect was not observed for hyperactivity/impulsivity or combined ADHD.[9] Finally, a large study (n=12,167) focusing on the relationship between eczema and ADHD, reported no effect of breastfeeding on ADHD after control for a range of confounding factors including SES.[10] Thus, studies investigating the protective effect of breastfeeding on AP/ADHD have shown inconclusive results. However, the effect of breastfeeding on ADHD is of particular interest in the light of a randomized controlled trial that suggested a hypersensitivity reaction to food as a causal factor for ADHD.[11] As breastfeeding has been suggested to protect against hypersensitivity reactions, breastfeeding might play a role in the development of AP/ADHD.

The positive effects of breastfeeding on brain development and IQ are thought to be mediated by the presence of long-chain polyunsaturated fatty acids (LC-PUFA) in breast milk, since LC-PUFA's such as docosahexanoic acid (DHA) and arachadonic acid (AA) play a role in neural function. Several studies have shown benefits of supplementation with LC-PUFA's or oily fish

intake during pregnancy for ADHD/AP and IQ.[12, 13] Although a meta-analysis on the effect of supplementation did not support a beneficial effect of LC-PUFA's on IQ,[14] another meta-analysis showed significant improvement of symptoms of ADHD.[15] DHA and AA are the products of a process in which essential fatty acids like 3-omega and 6-omega fatty acids are desaturated and elongated.[16, 17] The rate limiting step in the production of DHA and AA is mediated by the FADS enzymes. The importance of these fatty acids for neural development led Caspi et al. to study two SNPs in the FADS2 gene under the hypothesis that these SNPs could moderate the relationship between breastfeeding and IQ.[18] These two SNPs, rs174575 and rs1535, are in linkage disequilibrium (LD) with other SNPs throughout the promoter and intragenic region of the FADS2 gene. A gene-environment interaction was found in two independent cohorts; children carrying one or two C alleles of the rs174575 SNP showed a clear benefit of breastfeeding, whereas children homozygous for the G allele showed similar IQ scores in the breastfed and non-breastfed groups. Importantly, the interaction effect was significant after correction for maternal IQ and social class and it was shown that the interaction effect was not due to the maternal genotype influencing breast milk quality.[18]

Thus far, two studies have attempted to replicate this interaction effect. In a prospective cohort study of 6,000 children, the relationship between breastfeeding and IQ scores at age 8 was found to be modified by the two SNPs in the FADS2 gene.[19] However, the interaction effect was in the opposite direction of the effect observed by Caspi et al., with the effect of breastfeeding being significantly larger in children homozygous for the G allele.

A second replication study was performed in a cohort of 700 twin families that provided retrospective data on breastfeeding at age 16-18. Neither a main effect of breastfeeding nor an interaction of the FADS2 gene and breastfeeding was found after control for parental socioeconomic status and education.[20]

With the present study we contribute to the discussion whether breastfeeding is associated with IQ. We extend the IQ phenotype with associated phenotypes, namely educational attainment (EA) and overactive behavior (OA) / AP throughout childhood and investigate the possible role of FADS2 and its interaction with breastfeeding. Data on IQ, educational attainment (EA) and OA/AP were available for 1,739, 10,669 and 30,561 twins and siblings, respectively. Of these, 1,313, 1,857 and 2,849 individuals were successfully genotyped for the FADS2 SNPs.

Materials and methods

Subjects. The twins included in this study were registered as newborns with the Netherlands Twin Register (NTR).[21, 22] Longitudinal data on health and behavior are collected starting at registration. Questionnaires are sent out to the parents of the twins at registration and at twins' ages 2, 3, 5, 7, 10, and 12 years. Hereafter, their siblings are invited to participate as well, and both twins and siblings rate their own behavior. At age 18, the parents, twins and siblings are invited to participate in the adult register. Subsamples of twins and siblings are invited to participate in projects that use more elaborate measures such as IQ tests, MRI assessment or neuropsychological test to assess cognitive functioning and development.

IQ data were available for 1,739 individuals (birth cohorts 1974-1998), of whom 1,313 were genotyped for the FADS2 SNPs. IQ was measured in twins at ages 5, 7, 10, 12, and/or 18 and in siblings at age 12 and/or 18. Data on EA at age 12 were available for 10,669 twins and siblings (birth cohorts 1981-1999). Of these, 1,857 were genotyped for the FADS2 SNPs. Data for OA/AP at any time point were available for twins only (n=30,561, birth cohorts 1986-2004); 28,245 twins had OA data at age 3 and 18,296, 12,834, and 9,143 had data on AP at age 7, 10, and 12 years, respectively. Of these, 2,560, 2,479, 2,423, and 2,226 were genotyped for the FADS2 SNPs. Children suffering from a severe handicap that interfered with daily functioning were excluded from the analyses of OA/AP. Ethnic outliers were excluded from the analyses on the main and interaction effect of FADS2. If available, ethnicity was based on genome wide SNP data, otherwise information on the country of birth of the (grand)parents was used.

Breastfeeding. Breastfeeding status was reported in several surveys. For most twins, breastfeeding information was available from a survey administered at age 2 of the twins. Mothers reported the duration of breastfeeding for each child using the following categories: 'no', 'less than two weeks', '2 - 6 weeks', '6 weeks – 3 months', '3 – 6 months' and 'more than 6 months'. For all siblings and a small subgroup of twins, breastfeeding information was based on parental reports from three surveys of the adult register (in 1991, 1995 and 2009) when the children were age 5-33. In one of these surveys, two answer categories were used ('yes' and 'no'), while in the other questionnaires the same six answer categories were used as in the young twin register. Breastfeeding was re-coded into two categories: never breastfed ('no' and 'less than two weeks') versus ever breastfed (all other categories); as this dichotomization led to the

highest consistency across raters and across time in individuals measured on multiple occasions. When multiple reports were available, maternal reports were preferred over paternal reports; and the rating closest to the moment of breastfeeding was selected. In children with phenotypic data, the frequency of breastfeeding was 45.0% in twins and 71.4% in siblings. Twin pairs were concordant for breastfeeding in 97.8 % of the cases and twin-sibling pairs were concordant for breastfeeding in 72.8% of the cases. Of the children that were breastfed and had detailed information on the duration of breastfeeding, 40.6% were breastfed for more than three months (40.1% for twins and 63.5% for siblings). The frequency of breastfeeding was similar in the groups of genotyped and ungenotyped children (43.5% and 45.6%).

Maternal education. Maternal education was assessed in surveys sent out at twins' ages 3, 7 and 10, on a 13-point scale ranging from primary to post-doctoral education. The most recent measure of maternal education was re-coded into one of three categories: low, middle or high educational level. For a small subgroup, data on maternal education were based on surveys from the adult register and recoded into the same three categories as the data from the young register.

Zygosity. For twin pairs in which both twins were genotyped for the FADS2 gene, zygosity was based on a series of SNP markers and repeat polymorphisms or, where available, on genome-wide SNP data (van Beijsterveldt et al., 2012). In the twin pairs that had no FADS2 data, zygosity was determined by analysis of blood group or DNA polymorphisms in respectively 77.4%, 10.1% and 6.3% of same-sex twin pairs with data on IQ, EA and OA/AP. In the remaining cases, zygosity was based on opposite-sex information or a set of questions that gives a correct determination of zygosity in 93% of same-sex twin pairs.[23]

IQ. IQ was measured at ages 5, 7, 10, 12, and/or 18 as part of several studies.[6, 24-26] At ages 5, 7 and 10, children completed the Revised Amsterdamse Kinder Intelligentie Test (RAKIT).[27] The short version of the RAKIT was used, containing six subtests with age-appropriate items measuring verbal and nonverbal abilities. At age 10 and 12 the Dutch version of the Wechsler Intelligence Scale for Children-Revised (WISC-R and WISC-R-III) was used.[27-29] Both the complete test, consisting of 6 verbal and 6 nonverbal subtests and a short version, consisting of 6 subtests, were used. At age 18, IQ was assessed using the Dutch version of the Wechsler Adult Intelligence Scale (WAIS-III)[30] and the Raven Advanced Progressive Matrices.[31] All tests were standardized with equal norms across sex groups. Norms were based on a population sample of

same-aged subjects in the Netherlands, except for the Raven scores for which age-corrected standardized scores were calculated based on the NTR dataset. Standardized scores were then calculated for each project separately. For convenience, these z-scores were transformed to scores with mean 100 and standard deviation 15.

EA. The CITO-elementary test is a standardized test of Educational Attainment that is administered to 85% of Dutch children in their last year of primary education.[32] The test is taken on 3 consecutive days and covers four domains: Language, Mathematics, Information Processing and World Orientation. The total score ranges between 501 and 550. Bartels et al. showed that scores on the CITO-elementary test correlate .41, .50, .60, and .63 with IQ at ages 5, 7, 10, and 12, respectively.[33]

OA & AP. An age-appropriate version of the Child and Behavior Checklist (CBCL) was included in the questionnaires that were sent out to mothers of twins at ages 3, 7, 10 and 12 of the children. The Overactive scale at age 3 (OA) contains 5 items describing overactive behaviors and the Attention Problems scale (AP) at ages 7, 10 and 12 contains 11 items that describe both hyperactive and inattentive behaviours.[34-37] The sum score of the AP scale has been shown to converge with a DSM-based ADHD diagnosis.[38] As the two scales were analyzed simultaneously, scores were standardized by subtracting the mean score and dividing the outcome by the standard deviation.

FADS2 genotyping. The twins and siblings included in the study were genotyped as part of several projects. Most genotype data come from the SNP fingerprint chip that was used to determine zygosity and identify sample switches using a set of SNPs in candidate genes (see van Beijsterveldt et al., 2012). The remaining FADS2 genotype data come from genome-wide SNP arrays that were imputed against the 1000 genomes references set (June 2011, all panels) after stringent quality control. SNP quality control before imputation included filtering on the following criteria: HWE p-value > 0.00001; MAF > 0.01; SNP call rate > 0.95; SNP concordance rate < 2%; Mendel error rate < 2% and allele frequency difference with reference set < 0.20. C/G and A/T SNPs were only included if MAF < 0.35. Samples were included if the missing rate was < 10%, known gender was in line with the X-chromosome genotypes, IBS/IBD relationships were in line with known family relations and there were no issues of excessive heterozygosity or IBS sharing. The rs174575 and rs1535 SNPs were imputed with high quality

(R^2 for rs174575=0.93, R^2 for rs1535=0.98). Genotypes of the rs174575 and rs1535 SNPs were in Hardy-Weinberg equilibrium ($p > 0.05$) and allele frequencies were comparable to HapMap.

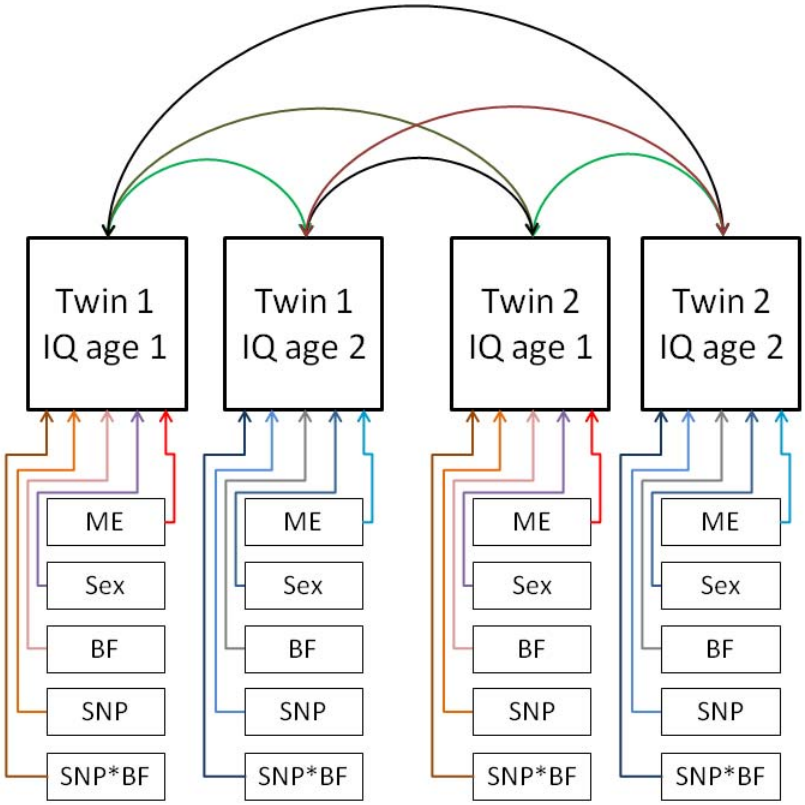
Analyses

To assess the effect of breastfeeding on IQ, EA and OA/AP, different structural equation models were fit; here, the EA, IQ and OA/AP scores were regressed on maternal education, sex and breastfeeding. To correct for the dependency in the data, covariances across ages and family members were estimated. For the analysis of EA, a single observation at age 12 was available for both twins and siblings, and a 3 x 3 matrix was estimated. In this matrix, the variance of EA and the covariances of EA within twin-sibling and twin pairs were estimated, with separate estimates for monozygotic and dizygotic twin pairs. For the analyses of IQ and OA/AP, longitudinal data were available, and we fit a model in which (co)variances were estimated across all available age groups. Thus, in the IQ analyses for instance, a 12 x 12 full covariance matrix was estimated (5 ages x 2 twins + 2 ages x 1 sibling). Separate covariance matrices were estimated for twin-sibling, monozygotic twin and dizygotic twin pairs, consisting of covariances between family members. The covariance across ages within individuals was constrained to be equal across twins and siblings.

The effects of breastfeeding, sex and maternal education were estimated as fixed effects in the total sample of genotyped and ungenotyped individuals. For IQ and OA/AP, it was first tested whether the effect of breastfeeding could be constrained to be equal across ages without a significant deterioration of model fit. If this was allowed it was tested whether the parameter could be excluded from the model. To assess the effect of FADS2 and the FADS2-by-breastfeeding interaction, similar models were fitted to the data of the subgroup with available rs1535 and rs174575 genotypes, with the SNP and SNP-by-breastfeeding effect as fixed effects in the model (see Figure 1). For IQ, it was first tested whether the breastfeeding, SNP and SNP-by-breastfeeding effect could be constrained to be equal across ages without a significant deterioration of model fit, both for each effect separately and for the breastfeeding, SNP and SNP-by-breastfeeding effects together. If this was allowed, it was tested whether excluding a predictor from the model across all ages led to a significantly worse model fit. For OA/AP, the effect of breastfeeding was significantly different across age and the breastfeeding and SNP-by-breastfeeding effect was therefore tested for each age separately. Predictors whose exclusion did

not lead to a significant decline in model fit where left out from subsequent models. The two SNPs were tested separately in an additive and a recessive model. The interaction of breastfeeding and FADS2 genotype was tested in the presence of main effects of breastfeeding and FADS2.

Figure 1. Model to estimate the effect of maternal education (ME), sex, breastfeeding (BF), SNP and SNP-by-breastfeeding (SNP*BF) on IQ. For the sake of clarity only two time points and two family members are shown. Included but not shown in the model are the within person residual variance of IQ and the intercept at each time point. Paths that have the same colour in the figure are constrained to be equal in the full model. Covariances across twins were estimated for monozygotic and dizygotic twins separately.



The Mx software package was used for all analyses.[39] Mx uses a full information maximum likelihood (FIML) method to fit a specified model to the data. The difference in minus two times the log-likelihood (-2LL) between two models has an asymptotic χ^2 distribution with the degrees of freedom (*df*) equalling the difference in parameters between the two models. The fit of nested

models can therefore be compared using the log-likelihood ratio test (LRT). As we analyzed three phenotypes, α-level of 0.05 was divided by the number of phenotypes: 0.05/3=0.017.
We performed a simulation study to estimate the power to detect the interaction effect described by Caspi et al. in the current dataset. The effect sizes for breastfeeding, rs174575 and its interaction were calculated based on the descriptive statistics reported by Caspi et al.; the interaction effect explained 0.3% of the variance in IQ in their dataset. We simulated ten thousand datasets that mimicked the structure of the current dataset with regard to sample size, family relations, longitudinal data structure and patterns of missingness, using the statistical package R (script available in Supplementary Material I). Effects were assumed to be equal across age and were not corrected for sex and maternal education, as only unadjusted effect sizes could be derived from the study by Caspi et al.

Results

The effect of breastfeeding on IQ was not significantly different across the different ages. After correction for sex and maternal education, the effect of breastfeeding was significant for EA at age 12 and marginally significant for IQ across all ages (p=0.05), see Table I. The effect of breastfeeding on OA/AP differed significantly across age and reached significance for OA at age 3. After correction for sex and maternal education, breastfed children scored on average 1.6 points (95% CI -0.00 - 3.14) higher on IQ, 1.3 points (95% CI 0.93-1.69) higher on the EA scale and 0.15 points (95% CI 0.09-0.21) lower on the OA scale.

For the children with genotype data, the mean IQ, EA and OA/AP scores for breastfed and non-breastfed children are shown in Table II, for each genotype group of rs174575 separately. This SNP showed the most convincing results for a gene-environment interaction in previous studies. In Figure 2 the same results are shown graphically for one measurement of each phenotype, OA at age 3 was selected because this was the only age for which a significant breastfeeding effect was found in the OA/AP analyses and IQ at age 18 was selected as this was the age at which most IQ data were available. Following previous studies, we first tested a recessive model. The main and interaction effects of breastfeeding and the child's genotype on IQ did not significantly differ across age, either when tested separately or with the three effects together. The main effect of the SNP and the SNP-by-breastfeeding interaction were not significant for IQ, EA and OA/AP for rs174575 and rs1535 (Table III). The point estimate of the interaction effect in the IQ analyses (not breastfed versus breastfed children coded 0-1 and CC/CG versus GG genotypes coded 0-1), was -5.30 (95% CI -11.32 - 0.73). The additive models also showed no significant results (Supplementary Table I). Results for the SNP and SNP-by-breastfeeding effect did not change meaningfully when not corrected for maternal education and sex (results available upon request).

The simulation study indicated that, at a significance level of 0.05, we were at 53% power to detect the previously reported interaction effect on IQ. For EA and OA/AP, we had 54% and 86% power to detect an interaction effect similar in effects size as the interaction effect on IQ reported by Caspi et al.

Table I. Model fitting results for the effect of breastfeeding in the total group with available data, adjusted for maternal education and sex.

	Model	vs Model	-2LL	df	X^2	Δdf	p	Conclusion
IQ	1. Full model		28605.199	3681				
	2. Equal β's BF across age	1	28605.963	3685	0.765	4	0.943	Effect of BF is the same across age
	3. Drop β breastfeeding	2	28609.804	3686	3.841	1	0.050	Effect of BF is marginally significant
EA	1. Full model		72965.872	10661				
	2. Drop β breastfeeding	1	73010.491	10662	44.619	1	< 0.001	Effect of BF is significant
OA/AP	1. Full model		165167.102	68472				
	2. Equal β's BF across age	1	165179.228	68475	12.126	3	0.007	Effect of BF is different at different ages
	3. Drop β BF age 3	1	165193.105	68473	26.003	1	< 0.001	Effect of BF is significant at age 3
	4. Drop β BF age 7	1	165167.368	68473	0.266	1	0.606	Effect of BF is not significant at age 7, 10 and 12
	5. Drop β BF age 10	4	165168.165	68474	0.797	1	0.372	
	6. Drop β BF age 12	5	165168.705	68475	0.540	1	0.462	

EA=educational attainment, OA=overactive behavior, AP=attention problems

Table II. Mean and standard deviation of IQ, educational attainment (EA), overactive behavior (OA) and attention problems (AP) scores for breastfed and non-breastfed children and genotype groups of rs174575 separately.

		rs174575 CC carriers				rs174575 CG carriers				rs174575 GG carriers			
		Not breastfed		Breastfed		Not breastfed		Breastfed		Not breastfed		Breastfed	
		n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)
IQ	Age 5	205	99.5 (14.9)	131	101.0 (13.8)	143	100.5 (14.9)	105	103.6 (14.4)	25	104.4 (10.7)	12	107.1 (10.7)
	Age 7	77	100.3 (14.3)	69	102.1 (14.5)	62	99.7 (14.2)	38	103.3 (15.1)	9	107.5 (12.3)	5	103.8 (8.9)
	Age 10	139	98.9 (13.6)	120	98.6 (15.4)	86	97.2 (14.3)	74	102.7 (13.5)	13	103.2 (13.2)	14	108.6 (10.7)
	Age 12	238	98.5 (14.0)	182	99.6 (14.5)	144	98.9 (13.9)	128	104.3 (14.4)	28	101.4 (13.1)	17	104.0 (12.9)
	Age 18	278	98.5 (15.5)	169	102.6 (12.9)	179	100.0 (15.2)	152	103.4 (13.6)	40	104.6 (13.4)	21	97.1 (13.9)
EA	Age 12	539	537.0 (8.2)	449	539.1 (8.4)	405	536.9 (8.8)	346	539.9 (8.0)	59	536.0 (9.3)	54	539.6 (8.1)
OA	Age 3	813	2.74 (2.18)	602	2.40 (2.14)	528	2.95 (2.22)	456	2.35 (2.10)	78	2.91 (2.53)	76	2.13 (1.81)
AP	Age 7	779	2.99 (3.00)	579	2.66 (2.89)	525	3.08 (2.99)	438	2.70 (3.10)	79	2.99 (3.28)	74	2.01 (2.03)
AP	Age 10	764	3.03 (3.08)	569	2.46 (2.87)	521	3.23 (3.34)	419	2.78 (3.37)	77	3.34 (3.28)	68	2.29 (2.44)
AP	Age 12	698	2.63 (2.77)	524	2.24 (2.80)	467	2.94 (3.14)	397	2.34 (2.71)	72	2.54 (3.30)	63	2.10 (1.73)

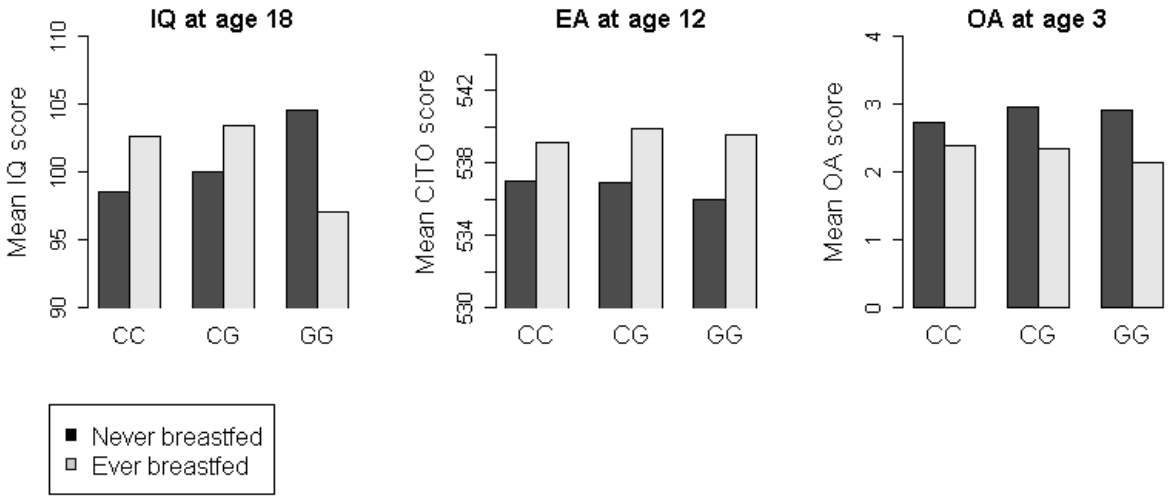
EA=educational attainment, OA=overactive behavior, AP=attention problems

Table III. Model fitting results for the recessive model of rs174575 and rs1535, adjusted for maternal education and sex.

			rs174575						rs1535				
	Model	vs	-2LL	df	X^2	Δdf	p		-2LL	df	X^2	Δdf	p
IQ	1. Full model		22070.409	2818					22172.311	2831			
	2. Equal β's across age	1	22079.993	2830	9.584	12	0.652		22185.362	2843	13.051	12	0.365
	3. Drop β SNP*BF	2	22082.960	2831	2.967	1	0.085		22188.678	2844	3.316	1	0.069
	4. Drop β SNP	3	22083.582	2832	0.622	1	0.430		22189.577	2845	0.899	1	0.343
EA	1. Full model		12547.612	1842					12569.755	1844			
	2. Drop β SNP*BF	1	12547.821	1843	0.209	1	0.648		12569.921	1845	0.166	1	0.684
	3. Drop β SNP	2	12548.077	1844	0.256	1	0.613		12572.235	1846	2.314	1	0.128
OA/AP	1. Full model		22092.026	9612					22134.252	9628			
	2. Drop β SNP*BF age 3	1	22092.974	9613	0.948	1	0.330		22134.603	9629	0.351	1	0.553
	3. Drop β SNP*BF age 7	2	22094.743	9614	1.769	1	0.184		22135.309	9630	0.706	1	0.401
	4. Drop β SNP*BF age 10	3	22094.774	9615	0.031	1	0.860		22135.969	9631	0.660	1	0.417
	5. Drop β SNP*BF age 12	4	22095.263	9616	0.489	1	0.484		22139.407	9632	3.438	1	0.064
	6. Equal β SNP across age	5	22097.483	9619	2.220	3	0.136		22147.392	9635	7.985	3	0.046
	7. Drop β SNP	6	22097.893	9620	0.410	1	0.522		22147.393	9636	0.001	1	0.975

EA=educational attainment, OA=overactive behavior, AP=attention problems

Figure 2. Mean IQ (age 18), educational attainment (EA, age 12) and overactive behavior (OA, age 3) scores, for breastfed and non-breastfed children and genotype groups of rs174575 separately.



Discussion

Breastfeeding was associated with educational attainment at age 12 and overactive behavior at age 3. For IQ at age 5-18, the effect was marginally significant after correction for maternal education. Polymorphisms in the FADS2 gene did not moderate the relationship between breastfeeding and IQ, educational attainment or overactive behavior. In addition, no main effects of the SNPs were found.

The positive effect of breastfeeding on IQ and EA is in line with findings of previous studies, including a study on the effect of breastfeeding on EA from our own group.[4, 5, 40] The effect of breastfeeding did not significantly differ across age, in line with a meta-analysis that found no moderation of the effect of breastfeeding by age at measurement.[41] A small protective effect of breastfeeding after correction for maternal education was found for OA at age 3, but not for AP at age 7, 10, and 12. Thus far, several studies on smaller samples suggested an effect of breastfeeding on ADHD/AP, but a large cohort study did not detect such an effect if the results were corrected for a range of confounding factors including SES.[7-10] The fact that we found a significant effect of breastfeeding only on OA at age 3 could suggest that the effect of breastfeeding is only present at a young age, or that the effect is specific to the OA scale. There are only 2 overlapping items between the OA and the AP scale; however, both scales describe hyperactive and inattentive behaviors with no clear difference in the overall content of the scales.

The results on the main effects of FADS2 on OA/AP are in line with a meta-analysis of genome-wide association studies on ADHD that showed no evidence for an association with any of the FADS2 SNPs included in the study.[42] Significant associations between FADS2 and AP/ADHD have only been reported for rs498793, which is in low LD with rs1535 and rs174575.[43]

The current study did not replicate the interaction effect between breastfeeding and the FADS2 genotype found by Caspi et al.[18] A previous replication effort reported an interaction effect in the opposite direction,[19] and another replication effort detected no interaction effect between FADS2 and breastfeeding on IQ.[20] Martin et al. stressed that the plausibility of the interaction effect of FADS2 and breastfeeding is not extensively supported given the fact that all studies to date have failed to detect a main effect of the FADS2 SNPs on IQ.[20] Munafo et al. showed that GxE interaction is unlikely to be present in the absence of a main effect of the genetic factor, given

sufficient power.[44] That is, when the environmental factor is not rare, the effect of the genotype in the exposed group will have a diminished but detectable effect in the total group of exposed and unexposed individuals and will thus be reflected as a significant main effect in the total group. However, it should be noted that the argument by Munafo et al. specifically focused on the case in which one of the groups (i.e. the unexposed group) shows no differences in the outcome variable across genotype groups. This is not the case for the results reported by Caspi et al.; breastfed children showed the lowest IQ scores in the GG group, whereas non-breastfed children showed the highest IQ scores in the GG group. This leads to the absence of a main effect of the genotype in the presence of a true interaction effect, as the genotype is advantageous in the non-breastfed group and non-advantageous in the breastfed group.

The sample size of the current study was smaller than the sample size of some of the previous studies. The simulation study showed that, at a significance level of 0.05, we were at 53%, 54%, and 86% power to detect an interaction effect on IQ, EA and OA/AP, respectively, assuming an interaction effect on IQ similar in size to the effect reported by Caspi et al. Altogether, the lack of an interaction effect in the present study and in the study by Martin et al. add to the conflicting results of the earlier studies.[18-20] The observed pattern is in line with the so called "winner's curse" phenomenon. In a context of low power in which claims of discovery are based on thresholds of statistical significance, newly discovered associations are prone to overestimate true effect sizes.[45] Further replication efforts and a successive meta-analysis are needed to further evaluate the possible interaction effect of FADS2 and breastfeeding on IQ.

Limitations. Participants included in the OA/AP analysis were born between 1986 and 2004. Modern formula contains DHA and AA since the early 2000s, which could lead to a smaller effect of breastfeeding on OA/AP in the later cohorts. While maternal education is a reasonable proxy for maternal IQ, the latter is to be preferred according to some researchers.[2] Although this could be regarded as a limitation for the analysis of the effect of breastfeeding, for the replication effort of the FADS2-by-breastfeeding effect this is of little importance, as the results by Caspi et al. were reported to be significant without correction for maternal IQ.

In conclusion, a protective effect of breastfeeding was found for overactive behavior at age 3 and educational attainment at age 12. These results were obtained after correction for maternal education. For IQ at ages 5, 7, 10, 12 and 18, the effect was marginally significant after

correction for maternal education. No main effect of SNPs in FADS2 nor an interaction between FADS2 SNPs and breastfeeding was detected for any of the phenotypes.

References

1. WHO. Global strategy for infant and young child feeding. Geneva: WHO; 2003.

2. Der G, Batty GD, Deary IJ. Effect of breast feeding on intelligence in children: prospective study, sibling pairs analysis, and meta-analysis. BMJ. 2006;333:945.

3. Jain A, Concato J, Leventhal JM. How good is the evidence linking breastfeeding and intelligence? Pediatrics. 2002;109:1044-1053.

4. Horta BL, Bahl R, Martines JC, Victoria CG. Evidence on the long-term effects of breastfeeding: systematic reviews and meta-analyses. Geneva: WHO; 2007.

5. Kramer MS, Aboud F, Mironova E, et al. Breastfeeding and child cognitive development: new evidence from a large randomized trial. Arch Gen Psychiatry. 2008;65:578-584.

6. Polderman TJ, Gosso MF, Posthuma D, et al. A longitudinal twin study on IQ, executive functioning, and attention problems during childhood and early adolescence. Acta Neurol Belg. 2006;106:191-207.

7. Kadziela-Olech H, Piotrowska-Jastrzebska J. The duration of breastfeeding and attention deficit hyperactivity disorder. Rocz Akad Med Bialymst. 2005;50:302-306.

8. Julvez J, Ribas-Fito N, Forns M, Garcia-Esteban R, Torrent M, Sunyer J. Attention behaviour and hyperactivity at age 4 and duration of breast-feeding. Acta Paediatr. 2007;96:842-847.

9. Al Hamed JH, Taha AZ, Sabra AA, Bella H. Attention Deficit Hyperactivity Disorder (ADHD) among male primary school children in Dammam, Saudi Arabia: prevalence and associated factors. J Egypt Public Health Assoc. 2008;83:165-182.

10. Romanos M, Gerlach M, Warnke A, Schmitt J. Association of attention-deficit/hyperactivity disorder and atopic eczema modified by sleep disturbance in a large population-based sample. J Epidemiol Community Health. 2010;64:269-273.

11. Pelsser LM, Frankena K, Toorman J, et al. Effects of a restricted elimination diet on the behaviour of children with attention-deficit hyperactivity disorder (INCA study): a randomised controlled trial. Lancet. 2011;377:494-503.

12. Gale CR, Robinson SM, Godfrey KM, Law CM, Schlotz W, O'Callaghan FJ. Oily fish intake during pregnancy--association with lower hyperactivity but not with higher full-scale IQ in offspring. J Child Psychol Psychiatry. 2008;49:1061-1068.

13. Hibbeln JR, Davis JM, Steer C, et al. Maternal seafood consumption in pregnancy and neurodevelopmental outcomes in childhood (ALSPAC study): an observational cohort study. Lancet. 2007;369:578-585.

14. Qawasmi A, Landeros-Weisenberger A, Leckman JF, Bloch MH. Meta-analysis of long-chain polyunsaturated fatty acid supplementation of formula and infant cognition. Pediatrics. 2012;129:1141-1149.

15. Bloch MH, Qawasmi A. Omega-3 fatty acid supplementation for the treatment of children with attention-deficit/hyperactivity disorder symptomatology: systematic review and meta-analysis. J Am Acad Child Adolesc Psychiatry. 2011;50:991-1000.

16. Lattka E, Illig T, Heinrich J, Koletzko B. Do FADS genotypes enhance our knowledge about fatty acid related phenotypes? Clin Nutr. 2010;29:277-287.

17. Lattka E, Illig T, Koletzko B, Heinrich J. Genetic variants of the FADS1 FADS2 gene cluster as related to essential fatty acid metabolism. Curr Opin Lipidol. 2010;21:64-69.

18. Caspi A, Williams B, Kim-Cohen J, et al. Moderation of breastfeeding effects on the IQ by genetic variation in fatty acid metabolism. Proc Natl Acad Sci U S A. 2007;104:18860-18865.

19. Steer CD, Davey Smith G, Emmett PM, Hibbeln JR, Golding J. FADS2 polymorphisms modify the effect of breastfeeding on child IQ. PLoS One. 2010;5:e11570.

20. Martin NW, Benyamin B, Hansell NK, et al. Cognitive function in adolescence: testing for interactions between breast-feeding and FADS2 polymorphisms. J Am Acad Child Adolesc Psychiatry. 2011;50:55-62 e54.

21. van Beijsterveldt CE, Groen-Blokhuis M, Hottenga JJ, et al. The Young Netherlands Twin Register (YNTR): longitudinal twin and family studies in over 70,000 children. Twin Res Hum Genet. 2013;16:252-267.

22. Willemsen G, Vink JM, Abdellaoui A, et al. The Adult Netherlands Twin Register: twenty-five years of survey and biological data collection. Twin Res Hum Genet. 2013;16:271-281.

23. Rietveld MJ, van Der Valk JC, Bongers IL, Stroet TM, Slagboom PE, Boomsma DI. Zygosity diagnosis in young twins by parental report. Twin Res. 2000;3:134-141.

24. Hoekstra RA, Bartels M, Boomsma DI. Longitudinal genetic study of verbal and nonverbal IQ from early childhood to young adulthood. Learn Individ Differ. 2007;17:97-114.

25. Rijsdijk FV, Vernon PA, Boomsma DI. Application of hierarchical genetic models to Raven and WAIS subtests: a Dutch twin study. Behav Genet. 2002;32:199-210.

26. van Soelen IL, Brouwer RM, van Leeuwen M, Kahn RS, Hulshoff Pol HE, Boomsma DI. Heritability of verbal and performance intelligence in a pediatric longitudinal sample. Twin Res Hum Genet. 2011;14:119-128.

27. Bleichrodt N, Drenth PJD, Zaal JN, Resing WCM. Revised Amsterdam Child Intelligence Test. Lisse, The Netherlands: Swets & Zeitlinger B.V.; 1984.

28. Van Haasen PP, De Bruyn EEJ, Pijl YJ, et al. Wechsler Intelligence Scale for Children-Revised, Dutch Version. Lisse, The Netherlands: Swets & Zetlinger B. V; 1986.

29. Wechsler D, Kort W, Compaan EL, Bleichrodt N, Resing WCM, Schittkatte M. Wechsler Intelligence Scale for Children-Third Edition. Lisse, The Netherlands: Swets and Zetlinger; 2002.

30. Wechsler D. Wechsler Adult Intelligence Scale–Third Edition Dutch version. Lisse, The Netherlands: Swets and Zetlinger; 1997.

31. Raven JC. Standard Progressive Matrices: Sets A, B, C, D, and E. : University Printing House, Cambridge. London: M.K. Lewis & Co; 1985.

32. CITO. Eindtoets basisonderwijs. Arnhem, the Netherlands: Citogroep; 2002.

33. Bartels M, Rietveld MJ, Van Baal GC, Boomsma DI. Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. Twin Res. 2002;5:544-553.

34. Achenbach TM. Manual for the Child Behavior Checklist/4-18 and 1991 Profile. Burlington, VT: University of Vermont, Department of Psychiatry; 1991.

35. Achenbach TM. Manual for the Child Behavior Checklist/2-3 and 1992 Profile. Burlington, VT: University of Vermont, Department of Psychiatry; 1992.

36. Koot HM, Van Den Oord EJ, Verhulst FC, Boomsma DI. Behavioral and emotional problems in young preschoolers: cross-cultural testing of the validity of the Child Behavior Checklist/2-3. J Abnorm Child Psychol. 1997;25:183-196.

37. Verhulst FC, Van der Ende, J., Koot, H. M. Handleiding voor de CBCL/4-18. Rotterdam: Afdeling Kinder- en Jeugdpsychiatrie, Sophia Kinderziekenhuis / Academisch Ziekenhuis Rotterdam / Erasmus Universiteit Rotterdam; 1996.

38. Derks EM, Hudziak JJ, Dolan CV, Ferdinand RF, Boomsma DI. The relations between DISC-IV DSM diagnoses of ADHD and multi-informant CBCL-AP syndrome scores. Compr Psychiatry. 2006;47:116-122.

39. Neale MC, Boker SM, Maes H. Mx Statistical modeling. 7th ed. Richmond, VA: Department of Psychiatry; 2006.

40. Bartels M, van Beijsterveldt CE, Boomsma DI. Breastfeeding, maternal education and cognitive function: a prospective study in twins. Behav Genet. 2009;39:616-622.

41. Anderson JW, Johnstone BM, Remley DT. Breast-feeding and cognitive development: a meta-analysis. Am J Clin Nutr. 1999;70:525-535.

42. Neale BM, Medland SE, Ripke S, et al. Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry. 2010;49:884-897.

43. Brookes KJ, Chen W, Xu X, Taylor E, Asherson P. Association of fatty acid desaturase genes with attention-deficit/hyperactivity disorder. Biol Psychiatry. 2006;60:1053-1061.

44. Munafo MR, Durrant C, Lewis G, Flint J. Gene X environment interactions at the serotonin transporter locus. Biol Psychiatry. 2009;65:211-219.

45. Ioannidis JP. Why Most Discovered True Associations Are Inflated. Epidemiology. 2008;19:640-648.

Supplementary Materials

Supplementary Table I. Model fitting results for the additive model of rs174575 and rs1535, adjusted for maternal education and sex.

			rs174575						rs1535				
	Model	vs	-2LL	df	X^2	Δdf	p		-2LL	df	X^2	Δdf	p
IQ	1. Full model		22075.891	2818					22176.344	2831			
	2. Equal β's BF across age	1	22082.093	2830	6.202	12	0.906		22186.431	2843	10.087	12	0.608
	3. Drop β SNP*BF	2	22082.131	2831	0.038	1	0.845		22189.110	2844	2.679	1	0.102
	4. Drop β SNP	3	22083.582	2832	1.451	1	0.228		22189.577	2845	0.467	1	0.494
EA	1. Full model		12547.841	1842					12571.455	1844			
	2. Drop β SNP*BF	1	12547.841	1843	0.000	1	0.626		12571.552	1845	0.097	1	0.755
	3. Drop β SNP	2	12548.078	1844	0.237	1	< 0.001		12572.235	1846	0.683	1	0.409
OA/AP	1. Full model		22086.988	9612					22135.234	9628			
	2. Drop β SNP*BF age 3	1	22089.555	9613	2.567	1	0.109		22135.900	9629	0.665	1	0.415
	3. Drop β SNP*BF age 7	2	22089.676	9614	0.121	1	0.728		22136.387	9630	0.487	1	0.485
	4. Drop β SNP*BF age 10	3	22090.060	9615	0.384	1	0.535		22139.061	9631	2.674	1	0.102
	5. Drop β SNP*BF age 12	4	22091.160	9616	1.100	1	0.294		22139.157	9632	0.096	1	0.757
	6. Equal β SNP across age	5	22096.625	9619	5.465	3	0.141		22146.332	9635	7.175	3	0.067
	7. Drop β SNP	6	22097.893	9620	1.268	1	0.260		22147.393	9636	1.061	1	0.303

EA=educational attainment, OA=overactive behavior, AP=attention problems

Supplementary Material I

R script to estimate the power to detect a gene (rs174575) by environment (breastfeeding/milk) effect in a dataset with related individuals with longitudinal measures. Due to the extensive missingness, the power is calculated by means of data simulation, i.e., we calculate the observed or empirical power based on the likelihood ratio test over many replications. To reduce computational burden, the background covariance matrix, i.e., conditional on the fixed effects, was fixed to its expected (true) value in optimizing the likelihood with and without the effects of interest (e.g., the interaction).

In the first part of the script several functions are created that are necessary to simulate and analyze the data, namely a loglikelihood function, a function to standardize data in a matrix and a multivariate normal density function.

Then, the number of datasets to simulate is specified as well as the data structure. The user can indicate the number of monozygotic (MZ) and dizygotic (DZ) twin families and the patterns of missingness that occur in the dataset (individuals from a family that have no data at all, and individuals that have no data at a particular time point).

Then, the MAF of the candidate gene, the frequency of breastfeeding and the correlation of breastfeeding across family members are specified as well as the effect sizes of the candidate gene, breast feeding and interaction effect expressed in terms of proportion of variance explained.

In addition, the alpha level and the background covariance matrices (covariance within person across time, covariance within MZ twin pairs, DZ twin pairs and twin-sibling pairs) can be specified.

Next, genotype and breastfeeding data are simulated. Then, phenotypic data are simulated given a particular genetic model (additive, dominant or recessive) and the effect sizes specified in the model. Subsequently, the simulated dataset is analyzed using the loglikelihood function and the statistical output is collected.

```
rm(list=ls(all=TRUE))
#
# semfitg is a function that calculates the derivatives of the loglikelihood
# function with respect to the parameters
# b0 (intercept), b1 (gene), b2 (milk), and b3 (gene x milk):
#
semfitg=function(par,alldatm,alldat,zyg,
nmz,ndz, eff, npar,smz,sdz) {
#
nfam=nmz+ndz
np=1 # parameter 4
gp=rep(0,1+sum(eff))
g0=g1=g2=g3=0
b0=par[np]
b1=b2=b3=0
if (eff[1]==1) {np=np+1; b1=par[np]}
if (eff[2]==1) {np=np+1; b2=par[np]}
if (eff[3]==1) {np=np+1; b3=par[np]}
#print(c(b1,b2,b3))
mus=matrix(0,nfam,22)
mus[,1:5]=(b1*alldat[,2]+b2*alldat[,5]+b3*alldat[,2]*alldat[,5])
mus[,6:10]=(b1*alldat[,3]+b2*alldat[,6]+b3*alldat[,3]*alldat[,6])
mus[,11:15]=(b1*alldat[,4]+b2*alldat[,7]+b3*alldat[,4]*alldat[,7])
mus=mus+b0
#
# derives ----------------------
for (i in 1:nfam) {
if (i==(ndz+1) | i==1) { test=rep(-1,15) }
cnew=(sum(test==alldatm[i,8:22]))==15
keep=which(alldatm[i,8:22]==1)
keep2=keep+7
n1=sum(alldatm[i,8:12]==1)
n2=sum(alldatm[i,13:17]==1)
n3=sum(alldatm[i,18:22]==1)
#
mum=as.vector(mus[i,keep])
if (zyg[i]==0 & !cnew) {
stmpi=solve(sdz[keep,keep]);
} #
if (zyg[i]==1 & !cnew) {
stmpi=solve(smz[keep,keep]) #
}
dldm=stmpi%*%as.matrix((alldat[i,keep2]-mum))
g0=g0-sum(dldm)
```

```
#
if (eff[1]==1) {
tmp=c(rep(alldat[i,2],n1),rep(alldat[i,3],n2),rep(alldat[i,4],n3))
g1=g1-sum(dldm*tmp)
}
if (eff[2]==1) {
tmp=c(rep(alldat[i,5],n1),rep(alldat[i,6],n2),rep(alldat[i,7],n3))
g2=g2-sum(dldm*tmp)
}
if (eff[3]==1) {
tmp=c(rep(alldat[i,5]*alldat[i,2],n1),
rep(alldat[i,6]*alldat[i,3],n2),
rep(alldat[i,7]*alldat[i,4],n3))
g3=g3-sum(dldm*tmp)
}
test=alldatm[i,8:22]
}
np=1
gp[np]=g0
if (eff[1]==1) {np=np+1; gp[np]=g1}
if (eff[2]==1) {np=np+1; gp[np]=g2}
if (eff[3]==1) {np=np+1; gp[np]=g3}
# print(gp)
gp
}
#
# The semfit function calculates the loglikelihood function
# We seek values of b0, b1, b2, b3 to maximize this function
# Once maximized the values of b0 b1 b2 b3 are maximumlikelihood estimates
#
semfit=function(par,alldatm,alldat,zyg,
nmz,ndz, eff, npar,smz,sdz) {
#
nfam=nmz+ndz
np=1 # parameter 4
b0=par[np]
b1=b2=b3=0
if (eff[1]==1) {np=np+1; b1=par[np]}
if (eff[2]==1) {np=np+1; b2=par[np]}
if (eff[3]==1) {np=np+1; b3=par[np]}
#print(c(b1,b2,b3))
mus=matrix(0,nfam,22)
mus[,1:5]=(b1*alldat[,2]+b2*alldat[,5]+b3*alldat[,2]*alldat[,5])
mus[,6:10]=(b1*alldat[,3]+b2*alldat[,6]+b3*alldat[,3]*alldat[,6])
mus[,11:15]=(b1*alldat[,4]+b2*alldat[,7]+b3*alldat[,4]*alldat[,7])
mus=mus+b0
```

```
#
# log likelihood function ----------------------
logl=0
for (i in 1:nfam) {
if (i==(ndz+1) | i==1) { tmp=rep(-1,15) }
cnew=(sum(tmp==alldatm[i,8:22]))==15
keep=which(alldatm[i,8:22]==1)
keep2=keep+7
mum=as.vector(mus[i,keep])
if (zyg[i]==0 & !cnew) {
#
sdzkk=as.matrix(sdz[keep,keep])
#
stmpi=solve(sdzkk);
#
#logdet=log(det(sdz[keep,keep]))
logdet=log(det(sdzkk)) # does not work no scalar s[i,i]
#
} #
if (zyg[i]==1 & !cnew) {
#
smzkk=as.matrix(smz[keep,keep])
#
stmpi=solve(smz[keep,keep]) #
#logdet=log(det(sdz[keep,keep]))
logdet=log(det(sdzkk))
#print(logdet)
}
#
# dmv is the multivariate normal density function
#
#if (length(keep2)>1) {
d=dmv(x=alldat[i,keep2],mean=mum,sigma=stmpi,logdet)
#}
#if (length(keep2)==1) {
#d=dnorm(x=alldat[i,keep2],mean=mum,sigma=sqrt(1/stmpi),log=T)
#}
# cannot happen
if (is.nan(d)) {return(100000)}
#
logl=logl+d
#
tmp=alldatm[i,8:22]
}
# print(c(logl,b0,b1,b2,b3))
#
```

```
-logl
}
#
# standi is a function to standardize data in a matrix
#
standi=function(dat) {
nr=dim(dat)[1]
nc=dim(dat)[2]
sds=diag(1/apply(dat,2,sd))
ms=t(as.matrix(apply(dat,2,mean)))
dat=(dat-(matrix(1,nr,1)%x%ms))%*%sds
dat
}
#
# dmv is the multivariate normal density function
#
dmv=function (x, mean, sigma, logdet)
{
  if (is.vector(x)) {
    x <- matrix(x, ncol = length(x))
  }
  distval <- mahalanobis(x, center = mean, cov = sigma, inverted=TRUE)
  logretval <- -(ncol(x) * log(2 * pi) + logdet + distval)/2
    return(logretval)
}
# ------------------------------------------------
#
# Start main program
# Specify the number of simulations, the data structure and patterns of missingness
#
library(MASS)
library(mvtnorm)
#
T=5                              # the number of measurements over time
nmem=3                           # the number of individuals per family
maxny=ny=T*nmem
#
# Specify the number of MZ and DZ twin families, the number of simulations and the
# patterns of missingness
#
selfseed=F
if (selfseed) set.seed(10208)
nrep=10000                               # specify the number of simulations
nmulti=1
nmz0=round(300*nmulti)                   # specify the number of monozygotic twin families
ndz0=round(344*nmulti)                   # specify the number of dizygotic twin families
```

```
#
nmismz=14     # specify the number of missingness patterns for monozygotic twin families
nmisdz=14     # specify the number of missingness patterns for dizygotic twin families
#
# now the patterns of missingness will be specified by use of a missing values
# indicator, 1=observed 0=missing. Each row specifies a particular pattern of
# missingness per family with the following columns: p geno milk t12345 t12345 t12345
#
#   p = proportion of the total dataset with this pattern of missingness
#   geno and milk indicate which individuals in the family have data available
#   geno = geno_twin1 geno_twin2 geno_sib
#   milk = milk_twin1 milk_twin2 milk_sib
#   t12345 indicates at which time points individuals have data available
#   t12345 t12345 t12345, IQ measurements 1-5 twin1, twin2, sib
#
missmz=matrix(c(
# p  geno  milk  t12345   t12345   t12345
.033, 1,0,0, 1,0,0, 1,1,1,1,1, 0,0,0,0,0, 0,0,0,0,0, # sample I –1 twin
.110, 1,1,0, 1,1,0, 1,1,1,1,1, 1,1,1,1,1, 0,0,0,0,0, # sample I -2 twins
.053, 1,1,1, 1,1,1, 1,1,1,1,1, 1,1,1,1,1, 0,0,0,0,1, # sample I –basis
.130, 1,1,0, 1,1,0, 0,0,1,1,0, 0,0,1,1,0, 0,0,0,0,0, # sample II -2 twins
.060, 1,0,0, 1,0,0, 1,0,0,1,0, 0,0,0,0,0, 0,0,0,0,0, # sample III –1 twin
.113, 1,1,0, 1,1,0, 1,0,0,1,1, 1,0,0,1,1, 0,0,0,0,0, # sample III –2 twins - a
.090, 1,1,0, 1,1,0, 1,0,0,1,0, 1,0,0,1,0, 0,0,0,0,0, # sample III –2 twins - b
.023, 1,1,1, 1,1,1, 1,0,0,1,1, 1,0,0,1,1, 0,0,0,1,0, # sample III –basis - a
.030, 1,1,1, 1,1,1, 1,0,0,1,0, 1,0,0,1,0, 0,0,0,1,0, # sample III –basis - b
.010, 1,0,0, 1,0,0, 0,0,0,0,1, 0,0,0,0,0, 0,0,0,0,0, # sample IV –1 twin
.157, 1,1,0, 1,1,0, 0,0,0,0,1, 0,0,0,0,1, 0,0,0,0,0, # sample IV -2 twins
.010, 1,0,0, 1,0,0, 0,0,0,0,1, 0,0,0,0,0, 0,0,0,0,0, # sample V -1 twin
.157, 1,1,0, 1,1,0, 0,0,0,0,1, 0,0,0,0,1, 0,0,0,0,0, # sample V -2 twins
.023, 1,1,1, 1,1,1, 0,0,0,0,1, 0,0,0,0,1, 0,0,0,0,1 # sample V -basis
),nmismz,22,byrow=T)
missdz=matrix(c(
# p  geno  milk  t12345   t12345   t12345
.012, 1,0,0, 1,0,0, 1,1,1,1,1, 0,0,0,0,0, 0,0,0,0,0, # sample I –1 twin
.160, 1,1,0, 1,1,0, 1,1,1,1,1, 1,1,1,1,1, 0,0,0,0,0, # sample I -2 twins
.073, 1,1,1, 1,1,1, 1,1,1,1,1, 1,1,1,1,1, 0,0,0,0,1, # sample I –basis
.148, 1,1,0, 1,1,0, 0,0,1,1,0, 0,0,1,1,0, 0,0,0,0,0, # sample II -2 twins
.015, 1,0,0, 1,0,0, 1,0,0,1,0, 0,0,0,0,0, 0,0,0,0,0, # sample III –1 twin
.076, 1,1,0, 1,1,0, 1,0,0,1,1, 1,0,0,1,1, 0,0,0,0,0, # sample III –2 twins - a
.105, 1,1,0, 1,1,0, 1,0,0,1,0, 1,0,0,1,0, 0,0,0,0,0, # sample III –2 twins - b
.061, 1,1,0, 1,1,0, 1,0,0,0,0, 1,0,0,0,0, 0,0,0,0,0, # sample III –2 twins - c
.023, 1,1,1, 1,1,1, 1,0,0,1,1, 1,0,0,1,1, 0,0,0,1,0, # sample III -basis
.049, 1,0,0, 1,0,0, 0,0,0,0,1, 0,0,0,0,0, 0,0,0,0,0, # sample IV –1 twin
.090, 1,1,0, 1,1,0, 0,0,0,0,1, 0,0,0,0,1, 0,0,0,0,0, # sample IV -2 twins
.020, 1,0,0, 1,0,0, 0,0,0,0,1, 0,0,0,0,0, 0,0,0,0,0, # sample V -1 twin
```

```
.137, 1,1,0, 1,1,0, 0,0,0,0,1, 0,0,0,0,1, 0,0,0,0,0, # sample V -2 twins
.032, 1,1,1, 1,1,1, 0,0,0,0,1, 0,0,0,0,1, 0,0,0,0,1 # sample V -basis
),nmisdz,22,byrow=T)
#
nmz=sum(round(missmz[,1]*nmz0)) # just in case ndz .ne. ndz0
ndz=sum(round(missdz[,1]*ndz0)) # just in case ndz .ne. ndz0
#
p=maf=.258                          # specify MAF of rs174575
pmilk=pm=.409                               # specify probability of being breastfed
#
# correlation of breastfeeding within family members.
# matrix: twin1 - twin2 - sib
#
dzmilk=mzmilk=matrix(c(
1,.997,.71,
.997,1,.71,
.71,.71,1),3,3,byrow=T)
#
# specify the effect sizes of the main effects and the interaction effect
# expressed as proportion of variance explained
#
mainmilk=0.04764                    # specify the effect size of breastfeeding
maincan=0.00002415                  # specify the effect size of rs174575
milkxcan=0.00265                    # specify the effect size of the interaction effect
#
alpha=.05                           #specify alpha
#
pu=matrix(0,nrep,5)
#
rtot=1-(mainmilk+maincan+milkxcan) # residual variance
#
b0=0 # intercept.
b1=sqrt(mainmilk)
b2=sqrt(maincan)
b3=sqrt(milkxcan)
# ------------------------------------------------
#
# specify the background covariance structure
#
# background correlations within the individual across time as observed in the
# current dataset.
#
Sph=matrix(c(
1.0000,0.5984,0.5782,0.5174,0.4599,
0.5984,1.0000,0.6915,0.6086,0.6446,
0.5782,0.6915,1.0000,0.7360,0.7216,
```

```
0.5174,0.6086,0.7360,1.0000,0.7181,
0.4599,0.6446,0.7216,0.7181,1.0000)*rtot,5,5)
#
# background correlations across dizygotic twins across time as observed in the
# current dataset.
#
Sdz2=matrix(c(
0.5487,0.3444,0.3204,0.2909,0.2338,
0.3444,0.4112,0.3282,0.3089,0.2440,
0.3204,0.3282,0.4336,0.4234,0.3297,
0.2909,0.3089,0.4234,0.4772,0.3738,
0.2338,0.2440,0.3297,0.3738,0.3632)*rtot,5,5)
#
# background correlations across monozygotic twins across time as observed in the
# current dataset.
#
Smz2=matrix(c(
0.6782,0.5711,0.5445,0.4899,0.4121,
0.5711,0.5541,0.6243,0.5557,0.4900,
0.5445,0.6243,0.7677,0.7167,0.6786,
0.4899,0.5557,0.7167,0.7998,0.6573,
0.4121,0.4900,0.6786,0.6573,0.7148)*rtot,5,5)
#
# background correlations across sibling-twin pairs across time as observed in the # current
dataset.
#
Ssib2=matrix(c(
0.2,0.2,0.2,0.2921,0.1424,
0.2,0.2,0.2,0.2,0.3453,
0.2,0.2,0.2,0.2,0.2440,
0.2,0.2,0.2,0.1569,0.1445,
0.2,0.2,0.2,0.1445,0.3488)*rtot,5,5)
#
# total 15 x 15 variance/covariance matrix in MZ families
#
Smz=rbind(
cbind(Sph,t(Smz2),t(Ssib2)),
cbind(Smz2,Sph,t(Ssib2)),
cbind(Ssib2,Ssib2,Sph))
#
# total 15 x 15 variance/covariance matrix in DZ families
#
Sdz=rbind(
cbind(Sph,t(Sdz2),t(Ssib2)),
cbind(Sdz2,Sph,t(Ssib2)),
cbind(Ssib2,Ssib2,Sph))
```

```
#
# now simulate genotypes for family members, start from parental genotypes
#
nmember=nmem2=nmem+2 # plus parents.
q=1-p
gf=c(p^2,2*p*q,q^2) # genotype freqs
pf=outer(gf,gf) # parent uncorrelated candidate.
#
nfamdz=ndz
nfammz=nmz
nfam=ndz+nmz #
nsib=nmem=3
nf=round(pf*nfam)
zyg=c(rep(0,nfamdz),rep(1,nfammz))
psib=array(0,c(3,3,3))
# A is decreasing its freq is MAF = p
#
# conditional offspring genotype frequencies
#
psib[1,1,1:3]=c(1,.0,.0) # parents AA=3 & AA=3, offspring AA
psib[1,2,1:3]=c(.5,.5,.0) # parents AA and Aa, offspring AA (.5) and Aa (.5)
psib[1,3,1:3]=c(.0,1,0) # parents AA and aa, offspring Aa
psib[2,1,1:3]=c(.5,.5,.0) # parents Aa and AA, offspring AA (.5) Aa (.5)
psib[2,2,1:3]=c(.25,.5,.25) # etc.
psib[2,3,1:3]=c(.0,.5,.5)
psib[3,1,1:3]=c(.0,1,.0)
psib[3,2,1:3]=c(.0,.5,.5)
psib[3,3,1:3]=c(.0,.0,1) # parents aa & aa, offspring aa
dose=c(3,2,1)
dose0=dose-1
dosep=c(1,2,3)
#
#specify the genetic model 'dom' or 'rec' or 'add'
#
gact='rec'
#
for (irep in 1:nrep) {
print(irep)
genos=matrix(0,nfam,nmem)
genosp=matrix(0,nfam,2)
#
# parental genotypes
#
genosp[,1]=sample(dosep,nfam,replace=T,prob=gf) # father
genosp[,2]=sample(dosep,nfam,replace=T,prob=gf) # mother
#
```

```
# offspring genotypes
#
for (i in 1:nfam) {
ig1=genosp[i,1] #*-1+4
ig2=genosp[i,2] #*-1+4
#
# conditional probabilities given parental genotypes
#
ptmp=psib[ig1,ig2,1:3]
#
# offspring genotypes
#
genos[i,1:nmem]=sample(dose,nsib,replace=T,prob=ptmp)
}
#
# mz twins are identical
#
for (i in 1:nfam) {
if (zyg[i]==1) {genos[i,2]=genos[i,1]}
}
#
# simulate the variable milk
#
milkdat=matrix(1,nfam,nmem)
tmp=mvrnorm(nfam,mu=rep(0,nmem),Sigma=dzmilk)
thresh=qnorm(pmilk)
milkdat[tmp>thresh]=0   # no milk
#
# collect the data in a matrix
#
alldat=matrix(0,nfam,22) # zyg m1-m3 g1-g3 t1-t5 t1-t5 t1-t5
#
alldat[,1]=zyg
alldat[,2:4]=milkdat
if (gact=='dom') genos[genos==3]=2 # 1,2=3
if (gact=='rec') { # 1=2, 3 recoded 1,2
genos[genos==2]=1;
genos[genos==3]=2;
}
alldat[,5:7]=genos[,1:3]-1 # coded 0,1
#
# background covariance structure
#
alldat[zyg==1,8:22]=mvrnorm(nmz,mu=rep(0,maxny),Sigma=Smz)
alldat[zyg==0,8:22]=mvrnorm(ndz,mu=rep(0,maxny),Sigma=Sdz)
#
```

```
# effect sizes main effects and milk X candidate gene
#
# ... b0+alldat[,2]*b1+alldat[,5]*b2+alldat[,2]*alldat[,5]*b3
#
alldat[,2:7]=standi(alldat[,2:7])
# add effects
alldat[,8:12]=alldat[,8:12]+
b0+alldat[,2]*b1+alldat[,5]*b2+alldat[,2]*alldat[,5]*b3
alldat[,13:17]=alldat[,13:17]+
b0+alldat[,3]*b1+alldat[,6]*b2+alldat[,3]*alldat[,6]*b3
alldat[,18:22]=alldat[,18:22]+
b0+alldat[,4]*b1+alldat[,7]*b2+alldat[,4]*alldat[,7]*b3
#
# now get missing indicators of each case
#
misnmz1=diffinv(round(missmz[,1]*nmz0))
misndz1=diffinv(round(missdz[,1]*ndz0)) + nmz
#
# the matrix for the missing indicator
#
alldatm=matrix(0,nfam,22) # zyg m1-m3 g1-g3 t1-t5 t1-t5 t1-t5
#
# introduce missingness
#
for (i in 1:nmismz) {
blockmz=round(missmz[i,1]*nmz0) # cvd 1 nov
tmp=matrix( missmz[i,2:22] , blockmz, 21, byrow=T)
alldatm[ (misnmz1[i]+1):(misnmz1[i+1]),2:22]=tmp
}
#
# missingness info is now in alldatm
#
for (i in 1:nmisdz) {
blockdz=round(missdz[i,1]*ndz0) # cvd 1 nov + correction maria
tmp=matrix( missdz[i,2:22] , blockdz, 21, byrow=T)
alldatm[(misndz1[i]+1):(misndz1[i+1]),2:22]=tmp
}
# now fit the model using maximum likelihood (ML) estimation and a fixed background
# covariance structure
#
par=c()
npar=1
par[npar]=b0 # bo
eff=c(1,1,1)
if (eff[1]==1) {npar=npar+1; par[npar]=b1} # b1 gene
if (eff[2]==1) {npar=npar+1; par[npar]=b2} # b2 milk
```

```
if (eff[3]==1) {npar=npar+1; par[npar]=b3} # b3 gene X milk
#
# ---------- ML estimation 4 parameters
#
fit1=T
fit2=T
pchi=-1
if (fit1) {
res1=optim(par,semfit,semfitg,method="L-BFGS-B",lower = -Inf, upper =
Inf,alldatm=alldatm,alldat=alldat,zyg=zyg,
nmz=nmz,ndz=ndz,eff=eff,npar=npar,smz=Smz,sdz=Sdz)
} # fit1
#
par=c()
npar=1
par[npar]=b0 # bo
eff=c(1,1,0)
if (eff[1]==1) {npar=npar+1; par[npar]=b1} # b1
if (eff[2]==1) {npar=npar+1; par[npar]=b2} # b2
if (eff[3]==1) {npar=npar+1; par[npar]=b3} # b3
#
# ---------- ML estimation 3 parameters drop b3
#
if (fit2) {
res2=optim(par,semfit,semfitg,method="L-BFGS-B",lower = -Inf, upper =
Inf,alldatm=alldatm,alldat=alldat,zyg=zyg,
nmz=nmz,ndz=ndz,eff=eff,npar=npar,smz=Smz,sdz=Sdz)
}
#
# likelihood ratio test of dropped parameter
#
if (fit1 & fit2) {
chi2=-2*(res1$value-res2$value)
pchi=pchisq(chi2,1,lower=F)
}
ures11=lm(alldat[,8]~alldat[,2]*alldat[,5],data=as.data.frame(alldat))
ures12=lm(alldat[,8]~alldat[,2]+alldat[,5],data=as.data.frame(alldat))
ures11vs2=anova(ures12,ures11)
ures21=lm(alldat[,13]~alldat[,3]*alldat[,6],data=as.data.frame(alldat))
ures22=lm(alldat[,13]~alldat[,3]+alldat[,6],data=as.data.frame(alldat))
ures21vs2=anova(ures22,ures21)
ures31=lm(alldat[,18]~alldat[,4]*alldat[,7],data=as.data.frame(alldat))
ures32=lm(alldat[,18]~alldat[,4]+alldat[,7],data=as.data.frame(alldat))
ures31vs2=anova(ures32,ures31)
#
# collect p values
```

```
#
pu[irep,1]=ures11vs2[6][2,1]
pu[irep,2]=ures21vs2[6][2,1]
pu[irep,3]=ures31vs2[6][2,1]
pu[irep,4]=pchi
pu[irep,5]=chi2
#
} # next replication
# show empirical power given alpha
#
epower=apply(pu[,1:4]<alpha,2,mean)
epower
```

Part II

Gene finding studies for ADHD and other childhood phenotypes

Chapter 5

Attention Deficit Hyperactivity Disorder polygenic risk scores predict Attention Problems in a population-based sample of children

Groen-Blokhuis MM, Middeldorp CM, Abdellaoui A, et al. Attention Deficit Hyperactivity Disorder polygenic risk scores predict Attention Problems in a population-based sample of children. J Am Acad Child Adolesc Psychiatry. Under review.

Abstract

Objective. Clinically, Attention Deficit Hyperactivity Disorder (ADHD) is characterized by hyperactivity, impulsivity and inattention and is among the most common childhood disorders. However, these same traits that define ADHD are variable in the general population and the clinical diagnosis may represent the extreme end of a continuous distribution of inattentive and hyperactive behaviors. This hypothesis can be tested by assessing the predictive value of polygenic risk scores derived from a discovery sample of ADHD cases in a target sample from the general population with continuous scores of inattention and hyperactivity. In addition, the genetic overlap between ADHD and continuous ADHD scores can be tested across rater and age.

Method. The Psychiatric Genomics Consortium has performed the largest GWA study of ADHD so far, including 5,621 clinical cases and 13,589 controls. The effects sizes of SNPs estimated in this meta-analysis were used to obtain individual polygenic risk scores in an independent population-based cohort of children from the Netherlands Twin Register. The variance explained in Attention Problems by the polygenic risk scores was estimated by linear mixed models.

Results. The ADHD polygenic risk scores significantly predicted both parent and teacher ratings of Attention Problems in preschool and school-age children.

Conclusions. These results indicate genetic overlap between a diagnosis of ADHD and Attention Problems scores across raters and age groups and provides evidence for a dimensional model of ADHD. Future GWA studies on ADHD can likely benefit from the inclusion of population-based cohorts and the analysis of continuous scores.

Introduction

Attention Deficit Hyperactivity Disorder (ADHD) is a condition characterized by age-inappropriate hyperactivity/impulsivity and inattention, resulting in significant impairment in about 5% of children.[1, 2] In the diagnostic manuals used in clinical practice, e.g. the ICD-10, DSM-IV and now the DSM-5,[1, 3, 4] a clinical diagnosis of ADHD is a binary trait which can be useful for defining treatment and care. However, ADHD might be the extreme end of a continuous distribution of inattentive and hyperactive behaviors that can be observed in the general population.[5-7] One approach to test for this dimensional model of ADHD is by assessing whether genetic risk factors for an ADHD diagnosis influence behavior across the entire spectrum of inattentive and hyperactive behavior. In addition, the evaluation of genetic factors for ADHD may clarify apparent differences in the ADHD assessment across raters and age groups; correlations between parent and teacher ratings are generally only moderate, as is the correlation within maternal ratings across preschool and school-age.[8-10] Previous twin studies support the validity of the dimensional model at the genetic level and indicate a large stability of genetic effects across both raters and time, but this has not yet been investigated using genome-wide genotype data.[8-12]

Both ADHD diagnosis and continuous measures of ADHD behaviors are highly heritable in childhood, with about 60-80% of the variance being due to genetic variation.[13-17] Despite this high heritability, current genome-wide association (GWA) studies have thus far been unsuccessful in detecting genetic risk variants for ADHD at genome-wide significant levels, suggesting a high degree of polygenic inheritance.[18] A study by the Psychiatric Genomics Consortium (PGC) showed that 28% of the liability to ADHD is explained by the SNPs present on current GWA platforms.[19] This implies indeed that many common variants of small effect stay undetected in current GWA studies due to limited sample size, but do contribute to the genetic liability of ADHD. The effect sizes obtained in ADHD GWA studies can be employed to estimate the genetic risk of the individual; so called polygenic risk scores are obtained by multiplying the observed number of risk alleles at a particular locus by the effect size observed in a GWA study summing over all SNPs that surpass a certain threshold of significance.[20, 21] With regard to ADHD specifically, polygenic risk scores based on the results of the PGC ADHD meta-analysis published in 2010, significantly predicted ADHD status in an independent sample

of 452 clinical ADHD cases and 5,081 controls, with higher polygenic risk scores in ADHD cases with comorbid aggression.[18, 22] Polygenic risk scores can also be used to assess the genetic overlap across traits. For example, polygenic risk scores based on a GWA study on schizophrenia have been utilized to test for the genetic overlap between schizophrenia and quantitative measures of psychosis.[23] Similarly, polygenic risk scores based on a GWA study in cases with major depressive disorder have been shown to be predictive of continuous scores of anxiety and depression in a general population sample.[24] In the current study, we obtained polygenic risk scores to assess the genetic overlap between ADHD as a clinical diagnosis and Attention Problems in a general population sample of children, using parent ratings at preschool age and parent and teacher ratings at school age.

Method

Genotype and phenotype data come from the Netherlands Twin Register (NTR) which was established in 1986.[25, 26] In the Young NTR (YNTR), surveys assessing the health and behavior of newborn twins are sent out to their parents at registration and at age 2, 3, 5, 7, 10 and 12 years. At age 7, 10 and 12, the teachers of the twins are invited to provide ratings of the children's behavior.

Attention Problems. Age-appropriate versions of the Achenbach System of Empirically Based Assessment (ASEBA) have been included in the YNTR surveys.[27, 28] At ages 3, 7, 10 and 12 the Child Behavior Checklist (CBCL) was collected from parents. At ages, 7, 10 and 12, the Teacher Report Form (TRF) has been included in teacher surveys. Respondents were asked to rate the child's behavior on ~120 items on a 3-point scale (0 = Not true; 1 = Somewhat or sometimes true; 2 = Very true or often true). The Attention Problems scale describes both hyperactive and inattentive behavior and contains 5 items in the CBCL at preschool age, 10 items in the CBCL at school-age and 26 items in the TRF at school-age. When multiple measures were available for the school-age (age 6-13) mother or teacher ratings, the measure closest to age 10 was chosen. As mother and father ratings were highly correlated (r = 0.71 and 0.73) and showed essentially similar results, we report only on the larger set of maternal ratings.

Genotype data. All children were genotyped on the Affymetrix 6.0 platform. Quality control and imputation were performed on a larger dataset that also included genotype data from the parents of the twins. Data were cleaned with the following criteria for SNP cleaning: Hardy-Weinberg equilibrium (HWE) p-value > 0.00001, minor allele frequency (MAF) > 0.01, call rate > 0.95, concordance rate in duplicate samples > 0.98, Mendelian error rate < 0.02 and allele frequency difference with reference set < 0.20. C/G and A/T SNPs were only included if MAF < 0.35. Samples were cleaned on the following criteria: call rate > 0.90, heterozygosity -0.10 < F < 0.10, consistency of X chromosome genotypes with known gender, consistency of expected and observed family relations between samples and Mendelian error rate < 0.02. Ethnic outliers were identified using principal component analysis and excluded from the analyses. Data were phased using Mach 1.0 and imputed with Minimac using all ethnicity panels of the 1000 Genomes Phase I Integrated Release Version 3 (2010-11-23 sequence data freeze, 2012-03-14 haplotypes).

Within PGC, a meta-analysis of 5,621 clinical ADHD cases and 13,589 controls has been conducted (Peter Holmans for the Psychiatrics Genomics Consortium (PGC), abstract of oral presentation 21st World Congress of Psychatric Genetics, October 17-21, Boston, USA,[19]). Polygenic risk scores were calculated in Plink. For each individual, the number of observed risk alleles at a particular locus (0,1,2) was multiplied with the ln(OR) observed in the PGC meta-analysis and summed over all SNPs. Several sets of polygenic risk scores were created based on different p-value thresholds in the discovery set (thresholds 0.0001, 0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5 and 1). SNPs were selected on the following criteria: info score > 0.30 in both the discovery and the target set, MAF > 0.02 in both sets, consistency of reported alleles across sets and a frequency difference across sets < 0.15. SNPs with C/G or A/T alleles were included only if MAF < 0.35.

Next, the Attention Problems scores were regressed on the polygenic risk scores in a linear mixed model. Polygenic risk scores were included as fixed effects in the model, while the dependence between measures in related individuals were accounted for by including an additive and a dominance genetic variance component. In all analyses, sex and age at measurement were included as covariates, as well as four Principal Components that reflect the Dutch population structure.[29] The polygenic risk scores and Attention Problems scales were standardized within each subset, that is, the mean of the subsample was subtracted from each score and divided by the standard deviation. The variance explained by the polygenic risk scores was then calculated by squaring the regression coefficient.

Results

Table 1 shows the number of individuals, and the means and standard deviations (SD) of age at measurement and Attention Problems scores for the mother and teacher ratings at preschool and school-age.

Table 1. Number of children and mean and standard deviation (SD) of age at measurement and Attention Problems scores for maternal ratings at preschool and school age, and for teacher ratings.

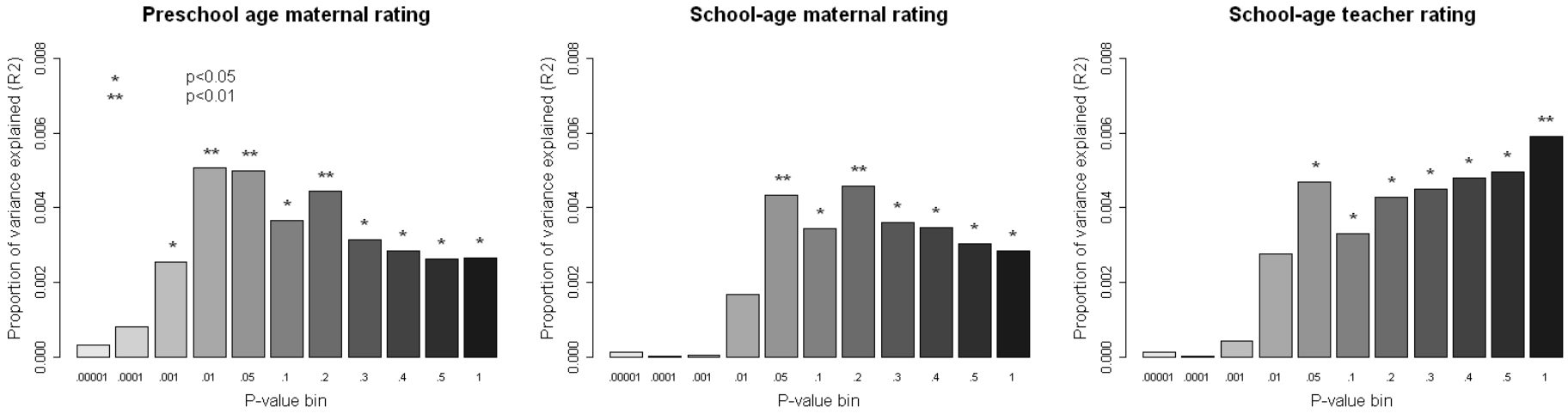
Measure	N	Age – mean (SD)	Attention Problems score – mean (SD)
Attention Problems preschool; age 3; mother rating CBCL	2,087	3.32 (0.26)	2.29 (2.00)
Attention Problems school; age 7-13; mother rating CBCL	2,132	9.93 (0.88)	3.26 (3.38)
Attention Problems school; age 7-13; teacher rating TRF	1,612	10.61 (1.40)	6.73 (7.92)

CBCL=Child Behavior Checklist, TRF=Teacher Report Form

Because of the longitudinal structure of the data, there is substantial overlap of individuals across these measures; 1,899 of 2,132 children with maternal ratings at school-age also had ratings at preschool age (phenotypic correlation 0.44), and 1,516 of 1,612 children with teacher ratings also had school-age maternal ratings (phenotypic correlation 0.49). In Figure 1, the variance in Attention Problems explained by the polygenic risk scores is shown, for the mother and teacher ratings at preschool and school-age, for different p-value bins. The polygenic risk scores based on the GWA meta-analysis results of clinical ADHD cases versus controls significantly predicted maternal ratings of Attention Problems at preschool and school-age, as well as teacher ratings at school-age. All significant effects were in the expected direction, i.e. higher polygenic risk scores were associated with higher Attention Problems scores. The maximum explained variance for each of the ratings varied between 0.5% and 0.6%, with no clear difference across preschool

and school-age, or mother and teacher ratings. All three measures showed significant skewness and kurtosis. To test whether the violation of distributional assumptions could influence the results analyses were repeated on quantile normalized scores (van der Waerden transformation[30], ranks averaged for tied data); results were essentially the same.

Figure 1. Proportion of variance explained in Attention Problems by polygenic risk scores based on different p-value bins. The number displayed below the bins is the upper threshold of p-values for inclusion, the lower theshold is always 0.



Discussion

Polygenic risk scores based on a GWA meta-analysis of clinical ADHD cases significantly predicted Attention Problems in an independent population-based sample of children. This indicates genetic overlap between ADHD assessed as a categorical disorder and Attention Problems assessed as a continuous trait in a general population sample. Other studies that examined whether ADHD should be considered a category or a continuum have used latent class analysis (LCA), factor analysis (FA), factor mixture models (FMM) and taxometric procedures on item level data to detect the underlying structure. Whereas some studies found evidence for distinct subtypes (e.g. predominantly inattentive, predominantly hyperactive/impulsive or combined inattentive and hyperactive/impulsive symptoms), others found evidence for the existence of a continuum within or across subtypes.[6, 31-35] The current study provides evidence in support of a dimensional model of ADHD at the genetic level. While categorical models of ADHD are necessary in clinical contexts where one needs to answer categorical questions with regard to e.g. treatment, future research on the underlying continuous trait could provide the necessary information to decide on the appropriate cutoffs for these categories. The National Institute of Mental Health has therefore launched the Research Domain Criteria (RDoC) project that aims to develop a research classification system for mental disorders based on dimensions of neurobiology and observable behavior.[36] Meanwhile, awareness of the dimensional nature of ADHD may already impact clinical practice, as shown by clinics that have successfully incorporated dimensional models in their daily clinical practice.[37]

Our finding also has important implications for gene finding studies. In line with previous studies, the results indicate that despite the lack of genome-wide significant findings, current GWA studies on ADHD contain a relevant signal.[19] Whereas one previous study drew on the genetic similarities across cases and controls based on SNP genotypes, the current study used the effect sizes from a PGC meta-analysis to predict Attention Problems in and independent sample. Another study that applied the same method to an independent sample of ADHD cases and controls also found a significant prediction of ADHD case status by ADHD GWA meta-analysis results.[22] These studies imply that GWA studies on ADHD can be successful, but that larger sample sizes are needed to identify common genetic variants for ADHD. Moreover, our results indicate that case-control cohorts can likely benefit from an increase in power by including the

available information on symptom severity in their analyses;[38] although this gain in power may be limited by the non-uniformity of measurement error across the distribution. Another possibility to increase power is the inclusion of super-normals as controls. In practice the choice of study design will likely depend on the costs of genotyping and phenotyping, and the availability of already existent datasets. In this context it is worthwhile to note that many population-based cohort studies have both genome-wide SNP data and continuous measures of ADHD available, but are currently underused for gene finding studies on ADHD and other psychiatric phenotypes. Although case-control studies benefit from the ascertainment of individuals from the extreme end of the distribution, the power to find genetic variants for ADHD is roughly equal in an equally sized population based cohort with a continuous measure of ADHD as in the latest PGC meta-analysis of ADHD, due to the relative high prevalence of ADHD and the somewhat small proportion of cases in the latest PGC meta-analysis. The latter follows from the formula for the NCP ratio between a case-control and population-based study as provided by Yang et al.[39]: $NCP_{01}/NCP_{QT} = (i^2v(1-v)N_{01})/(1-K)^2N_{QT}$, where K is the disease prevalence, N_{01} is the sample size of the case-control study, N_{QT} is the sample size of the population based study, v is the proportion of cases in the case-control study and i=z/k, with z the height of the standard normal distribution at the threshold of disease prevalence (if for ADHD a prevalence of 0.05 is assumed, i=2.063). Other advantages of population based studies include the richness of available phenotypic information allowing for multivariate analyses and the investigation of gene-environment interactions.[40]

Polygenic risk scores were predictive of Attention Problems at both preschool and school-age. This similarity of genetic effects across time is in line with twin studies that demonstrate a high genetic stability of maternal ratings of Attention Problems in childhood.[10, 11] A similar pattern of explained variance was observed for mother and teacher ratings of Attention Problems at school age. This is in agreement with the literature; although the phenotypic correlation of maternal and teacher ratings of Attention Problems is usually only moderate, the behavior that parent and teachers rate in common is highly heritable.[8, 41] Moreover, an ADHD diagnosis requires the behavior to be present in multiple settings, it is therefore expected that the genetic factors that influence a clinical diagnosis of ADHD correlate with both parent and teacher ratings of ADHD symptoms. Nevertheless, our finding that the genetic overlap with ADHD extends to both mother and teacher ratings of Attention Problems is informative for GWA studies as it indicates that

both ratings can be included, and that even ratings as early in childhood as age 3 contain information important to GWA studies. Eventually, lumping of all available data irrespective of rater and age could be considered for genetic studies. The resulting increase in sample size and statistical power probably outweighs the negative effects of increased phenotypic heterogeneity. In any case, possible age and rater-specific genetic effects are likely of less interest than genetic effects that influence ADHD scores across raters and age groups.

The assessment of genetic overlap by means of polygenic risk scores has some limitations. First, the genetic overlap that can be assessed is restricted to the set of SNPs that are present in both the discovery and target set, which effectively limited the analyses in the current study to SNPs included in HapMap with MAF > 0.02. Second, although the discovery set that was used is the largest ADHD GWA meta-analysis available, its sample size still limits the accuracy of the estimates of the small effect sizes typically observed in psychiatric GWA studies. It is therefore not surprising that the total variance of Attention Problems explained by the polygenic risk scores was below 1%. As a comparison, a study of 452 clinical ADHD cases and 5,081 controls in which polygenic risk scores were calculated based on a previous PGC ADHD meta-analysis (based on 2,064 trios, 896 case subjects and 2,455 comparisons) reported 0.098% explained variance.[18, 22] Although these estimates are not directly comparable due to difficulties with the interpretation of Nagelkerke's R^2 and the use of different discovery sets, it does show that the current observed variance explained is not exceptional.[42, 43] Finally, we included all SNPs that surpassed a certain threshold of significance in our polygenic risk scores, thereby ignoring the fact that SNPs in close proximity are often not in linkage disequilibrium and thus do not provide an independent association signal. When the analyses were repeated with a clumping procedure in Plink (option --clump, with settings --clump-p1 1 --clump-p2 1 --clump-r2 0.25 --clump-kb 500) the explained variance was somewhat lower for preschool maternal and school teacher ratings and non-significant for the maternal ratings at school age (results available upon request from the first author).

As both our and the previous study are able to predict ADHD/Attention Problems from ADHD polygenic risk scores, it is clear that the PGC ADHD meta-analysis picks up genetic variation relevant to ADHD, encouraging current efforts to increase the sample size for ADHD GWA studies. In the mean time, the available GWA results can be used to investigate the underlying

structure of ADHD and other relevant issues, such as the genetic overlap between ADHD and other disorders, rater and age effects, and gene-environment interplay. Our study supports the use of dimensional models of ADHD and indicates that future GWA studies can benefit from the inclusion of both population-based and case-control studies, and by analyzing ADHD as a quantitative rather than a categorical trait.

References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Arlington, VA: American Psychiatric Publishing; 2013.

2. Polanczyk G, de Lima MS, Horta BL, Biederman J, Rohde LA. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. Am J Psychiatry. 2007;164:942-948.

3. World Health Organisation. ICD-10 Classifications of Mental and Behavioural Disorder: Clinical Descriptions and Diagnostic Guidelines. Geneva: World Health Organisation; 1992.

4. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed, text revision. Washington, DC: American Psychiatric Association; 2000.

5. Hudziak JJ, Achenbach TM, Althoff RR, Pine DS. A dimensional approach to developmental psychopathology. Int J Methods Psychiatr Res. 2007;16 Suppl 1:S16-23.

6. Lubke GH, Hudziak JJ, Derks EM, van Bijsterveldt TC, Boomsma DI. Maternal ratings of attention problems in ADHD: evidence for the existence of a continuum. J Am Acad Child Adolesc Psychiatry. 2009;48:1085-1093.

7. Pickles A, Angold A. Natural categories or fundamental dimensions: on carving nature at the joints and the rearticulation of psychopathology. Dev Psychopathol. 2003;15:529-551.

8. Derks EM, Hudziak JJ, Beijsterveldt CE, Dolan CV, Boomsma DI. Genetic analyses of maternal and teacher ratings on attention problems in 7-year-old Dutch twins. Behav Genet. 2006;36:833-844.

9. Kuntsi J, Rijsdijk F, Ronald A, Asherson P, Plomin R. Genetic influences on the stability of attention-deficit/hyperactivity disorder symptoms from early to middle childhood. Biol Psychiatry. 2005;57:647-654.

10. Rietveld MJ, Hudziak JJ, Bartels M, van Beijsterveldt CE, Boomsma DI. Heritability of attention problems in children: longitudinal results from a study of twins, age 3 to 12. J Child Psychol Psychiatry. 2004;45:577-588.

11. Kan K, Dolan CV, Nivard MG, et al. Genetic and environmental stability in attention problems across the lifespan: evidence from the Netherlands twin register. J Am Acad Child Adolesc Psychiatry. 2013;52:12-25.

12. Larsson H, Anckarsater H, Rastam M, Chang Z, Lichtenstein P. Childhood attention-deficit hyperactivity disorder as an extreme of a continuous trait: a quantitative genetic study of 8,500 twin pairs. J Child Psychol Psychiatry. 2012;53:73-80.

13. Eaves LJ, Silberg JL, Meyer JM, et al. Genetics and developmental psychopathology: 2. The main effects of genes and environment on behavioral problems in the Virginia Twin Study of Adolescent Behavioral Development. J Child Psychol Psychiatry. 1997;38:965-980.

14. Gjone H, Stevenson J, Sundet JM. Genetic influence on parent-reported attention-related problems in a Norwegian general population twin sample. J Am Acad Child Adolesc Psychiatry. 1996;35:588-596.

15. Hudziak JJ, Rudiger LP, Neale MC, Heath AC, Todd RD. A twin study of inattentive, aggressive, and anxious/depressed behaviors. J Am Acad Child Adolesc Psychiatry. 2000;39:469-476.

16. Rietveld MJ, Hudziak JJ, Bartels M, van Beijsterveldt CE, Boomsma DI. Heritability of attention problems in children: I. cross-sectional results from a study of twins, age 3-12 years. Am J Med Genet B Neuropsychiatr Genet. 2003;117B:102-113.

17. Sherman DK, McGue MK, Iacono WG. Twin concordance for attention deficit hyperactivity disorder: a comparison of teachers' and mothers' reports. Am J Psychiatry. 1997;154:532-535.

18. Neale BM, Medland SE, Ripke S, et al. Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry. 2010;49:884-897.

19. Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee SH, Ripke S, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet. 2013;45:984-994.

20. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009;460:748-752.

21. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from SNPs. Nat Rev Genet. 2013;14:507-515.

22. Hamshere ML, Langley K, Martin J, et al. High loading of polygenic risk for ADHD in children with comorbid aggression. Am J Psychiatry. 2013;170:909-916.

23. Derks EM, Vorstman JA, Ripke S, Kahn RS, Ophoff RA. Investigation of the genetic association between quantitative measures of psychosis and schizophrenia: a polygenic risk score analysis. PLoS ONE. 2012;7:e37852.

24. Demirkan A, Penninx BW, Hek K, et al. Genetic risk profiles for depression and anxiety in adult and elderly cohorts. Mol Psychiatry. 2011;16:773-783.

25. van Beijsterveldt CE, Groen-Blokhuis M, Hottenga JJ, et al. The Young Netherlands Twin Register (YNTR): longitudinal twin and family studies in over 70,000 children. Twin Res Hum Genet. 2013;16:252-267.

26. Willemsen G, Vink JM, Abdellaoui A, et al. The Adult Netherlands Twin Register: twenty-five years of survey and biological data collection. Twin Res Hum Genet. 2013;16:271-281.

27. Achenbach TM, Rescorla LA. Manual for the ASEBA Preschool forms & Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families; 2000.

28. Achenbach TM, Rescorla LA. Manual for the ASEBA School-Age Forms & Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families; 2001.

29. Abdellaoui A, Hottenga JJ, de Knijff P, et al. Population structure, migration, and diversifying selection in the Netherlands. Eur J Hum Genet. 2013;21:1277-1285.

30. Lehmann EL. Nonparametrics: statistical methods based on ranks. San Francisco: Holden-Day; 1975.

31. Frazier TW, Youngstrom EA, Naugle RI. The latent structure of attention-deficit/hyperactivity disorder in a clinic-referred sample. Neuropsychology. 2007;21:45-64.

32. Hudziak JJ, Heath AC, Madden PF, et al. Latent class and factor analysis of DSM-IV ADHD: a twin study of female adolescents. J Am Acad Child Adolesc Psychiatry. 1998;37:848-857.

33. Lubke GH, Muthen B, Moilanen IK, et al. Subtypes versus severity differences in attention-deficit/hyperactivity disorder in the Northern Finnish Birth Cohort. J Am Acad Child Adolesc Psychiatry. 2007;46:1584-1593.

34. Neuman RJ, Todd RD, Heath AC, et al. Evaluation of ADHD typology in three contrasting samples: a latent class approach. J Am Acad Child Adolesc Psychiatry. 1999;38:25-33.

35. Todd RD, Rasmussen ER, Neuman RJ, et al. Familiality and heritability of subtypes of attention deficit hyperactivity disorder in a population sample of adolescent female twins. Am J Psychiatry. 2001;158:1891-1898.

36. Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. BMC Med. 2013;11:126.

37. Rettew D. Child temperament: new thinking about the boundary between traits and illness. New York: W.W. Norton & Company; 2013.

38. Lee SH, Wray NR. Novel genetic analysis for case-control genome-wide association studies: quantification of power and genomic prediction accuracy. PLoS One. 2013;8:e71494.

39. Yang J, Wray NR, Visscher PM. Comparing apples and oranges: equating the power of case-control and quantitative trait association studies. Genet Epidemiol. 2010;34:254-257.

40. Plomin R, Haworth CM, Davis OS. Common disorders are quantitative traits. Nat Rev Genet. 2009;10:872-878.

41. Thapar A, Harrington R, Ross K, McGuffin P. Does the definition of ADHD affect heritability? J Am Acad Child Adolesc Psychiatry. 2000;39:1528-1536.

42. Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 2013;9:e1003348.

43. Lee SH, Goddard ME, Wray NR, Visscher PM. A better coefficient of determination for genetic profile analysis. Genet Epidemiol. 2012;36:214-224.

Chapter 6

A GWA meta-analysis of continuous measures of ADHD symptoms in nine population based pediatric cohorts

Groen-Blokhuis MM*, Hammerschlag AR*, Ouwens KG, Steinberg J, St. Pourcain B, Trzaskowski M, Pappa I, van der Most P, Tiesler C, Ang W, Vilor N, Bacelis J, Greven C, Wang C, Jugessur A, Marinelli M, Hottenga JJ, Sengpiel V, Bustamante M, Jacobsson B, Whitehouse A, Nolte I, Ormel J, Heinrich J, Pennell C, EArly Genetics & Lifecourse Epidemiology (EAGLE) Consortium, Neale B, Plomin R, Tiemeier H, Hudziak JJ, Middeldorp CM**, Boomsma DI**

* these authors contributed equally to this work

** these authors jointly directed this work

Abstract

Within the field of childhood psychiatric genetics, the first GWA studies have been initiated. One of the phenotypes of interest is Attention Deficit Hyperactivity Disorder (ADHD) which has a heritability of around 75 percent. As of yet, no genome-wide significant findings have been reported. Previous studies have nevertheless shown that common SNPs explain a substantial part of the liability to ADHD and that there is a genetic overlap between the ADHD diagnosis and continuous measures of ADHD symptoms in the general population. Therefore, continuous measures of ADHD symptoms available in population-based cohorts can be used in gene finding studies for ADHD.

Within the EAGLE consortium, nine population-based cohorts including a total of 17,560 children had ADHD symptom data available as well as genotype data imputed against the 1000 Genomes reference panel. ADHD symptoms were rated by mothers and teachers at preschool and school age. Each cohort performed a linear regression of the symptom score on genotype dosages, sex, age and principal components. A p-value based meta-analysis was performed; the results of this meta-analysis were taken further and tested for gene-based associations and enrichment of biological pathways, using both biological processes and cellular annotations and information on knock-out mouse phenotypes.

None of the variants included in the GWA meta-analysis reached genome-wide significance, and no genes or pathways were associated with ADHD symptoms at a false-discovery rate of 5%. GWA studies of larger samples will thus be needed to detect genetic variants for ADHD related traits; the inclusion of population-based cohorts in GWA studies of ADHD can help to increase sample size and hence improve statistical power for gene finding.

Introduction

Attention Deficit Hyperactivity Disorder (ADHD) is a common psychiatric condition in childhood; its prevalence is around five percent across countries worldwide.[1] As an objective test is lacking, a diagnosis is made by a clinician based on the occurrence of age-inappropriate impulsive, hyperactive and inattentive behaviors that occur in multiple settings and cause significant impairment.[2, 3] The heritability of ADHD is estimated around 75 percent, yet no genetic variants have been identified that show a replicable association with ADHD.[4, 5] The largest genome-wide association (GWA) study so far has been performed by the Psychiatric Genomics Consortium (PGC) and included 5,621 cases and 13,589 controls. No SNPs exceeded the threshold of genome-wide significance.[6] In a sample that overlapped this meta-analysis (but excluding a Chinese case-control study), a genetic complex trait analysis (GCTA) was performed, in which the variance explained by all SNPs is calculated by comparing the phenotypic similarity to the observed genotypic similarity across all possible pairs included in the study. It was estimated that 28% of the liability to ADHD can be attributed to SNPs included in GWA studies.[7] This suggests that GWA studies can be effective in identifying genetic risk variants for ADHD, but that current samples lack the statistical power to do so.

Besides increasing sample sizes, power in GWA studies can be enhanced by improving phenotypic measurements.[8] Several studies have indicated that ADHD can be regarded the extreme end of a continuous distribution of inattentive and hyperactive behaviors.[9-11] Genetic epidemiological studies indicate that there is a substantial overlap between the genetic factors for a clinical diagnosis of ADHD and continuous measures of ADHD in the general population.[10, 11] Moreover, a molecular genetic study showed that the effect sizes reported in the latest PGC ADHD GWA could be used to predict Attention Problems in an independent population-based cohort of children, suggesting that the genetic overlap is present for variants included in GWA studies.[12]

Although case-control studies benefit from the oversampling of the high scoring end of the distribution, using the full information on symptom severity in the population can enhance power, especially for a relatively common disorder as ADHD.[8, 13] Many population-based cohorts have genome-wide SNP data and continuous scores of ADHD symptoms available and

thus provide an excellent yet underused opportunity for gene-finding studies for ADHD. The current report describes the first GWA meta-analysis of continuous measures of ADHD symptoms in population-based samples of children including a total of nine cohorts with a total sample size of 17,560 children.

Materials and Methods

GWA meta-analysis. The EArly Genetics & Lifecourse Epidemiology (EAGLE) consortium is a collaboration between several population based birth cohorts from Europe, Australia and the United States. The consortium focuses on a wide range of phenotypes in childhood; previous EAGLE studies have e.g. reported on GWA studies of eczema, infant head circumference, and internalizing problems.[14-16] For the current study, all EAGLE cohorts with continuous measures of ADHD-like behaviors in childhood (age at measurement < 13 years) were invited to participate in the meta-analysis. In addition, several other pediatric cohorts of Caucasian descent with genotype and ADHD symptom data were approached to participate in the current project. Different measurement instruments were used across cohorts, such as the Attention Problems scale of the Child Behavior Checklist (CBCL), the Hyperactivity scale of the Strengths and Difficulties Questionnaire (SDQ) and the DSM-IV ADHD items as included in the Conners' Rating Scale (see Supplementary Table 1 for the items included in each scale).[3, 17-20] If available, cohorts were asked to analyze the scores of different measurement instruments, preschool (ages 0 through 5) and school-age (ages 6 through 12) ratings, and mother and teacher ratings separately. For the meta-analysis, one rating was selected for each cohort: where applicable, school-age ratings were preferred over preschool age ratings, mother ratings were preferred over teacher ratings and the measurement instrument with the largest information density was preferred over the other instruments (DSM-IV > CBCL > SDQ). An overview of the cohorts and the ADHD measures included in the meta-analysis is provided in Table 1. All cohorts recruited children during pregnancy or at birth, except for TRAILS where children were recruited at age 10-12. More information on the individual cohorts can be found on the websites and in the papers listed in Table 1.

Table 1. Description of the cohorts and ADHD symptom scores included in the meta-analysis.

Cohort	N	Measurement instrument	Rater	Age in years mean (SD)	Sum score mean (SD)	Website	Reference papers
ALSPAC	5,757	SDQ	Parent	9.65 (0.12)	2.91 (2.24)	www.bristol.ac.uk/alspac/	[21]
Generation R	2,211	CBCL/1.5-5	Parent	6.01 (0.38)	1.38 (1.69)	www.generationr.nl	[22]
GINI / LISA	1,389	SDQ	Parent	10.04 (0.20)	2.71 (2.36)	www.helmholtz-muenchen.de/epi/arbeitsgruppen/umweltepidemiologie/projects-projekte/lisa-plus/index.html	[23]
INMA	804	DSM	Teacher	4.91 (0.69)	5.38 (6.83)	www.proyectoinma.org/	[24]
MOBA	665	CBCL/1.5-5	Parent	3.05 (0.10)	2.05 (1.67)	www.fhi.no/morogbarn	[25]
NTR	1,605	CBCL/6-12	Parent	9.95 (0.85)	3.25 (3.39)	www.tweelingenregister.org	[26]
Raine	1,338	CBCL/6-12	Parent	10.58 (0.20)	2.60 (3.17)	www.rainestudy.org.au	[27-29]
TEDS	2,606	Conners'	Parent	7.88 (0.52)	10.51 (8.62)	www.teds.ac.uk	[30]
TRAILS	1,285	CBCL/6-12	Parent	11.08 (0.56)	4.27 (3.40)	www.trails.nl	[31]

All cohorts imputed their genotype data with the March 2012 release of the 1000 Genomes reference set using all ethnicity panels. Cohorts were asked to perform a linear regression of the sum score of the ADHD symptom scale on sex, age at measurement, genotype dose, and principal components as necessary. All cohorts analyzed sets of unrelated individuals, except for the NTR that included two individuals of each dizygotic twin pair; in this cohort standard errors were corrected in Plink with the --family option. An overview of the imputation and analysis procedure followed in each cohort is provided in Supplementary Table 1. Results were checked and meta-analyzed by two independent analysts. Quality control included calculation of the inflation factor lambda (calculated as the observed median chi-square divided by the expected median chi-square), format checking, visual inspection of QQ plots, Manhattan plots, and histograms of minor allele frequency (MAF) and INFO scores, consistency of reported allele frequency with the reference set (1000 Genomes, EUR panels), consistency of reported p-value with reported beta and standard error (SE) and consistency of reported SE with reported sample size, SD and MAF. All files were filtered based on INFO metric > 0.3, expected minor allele count (EMAC; 2*N*MAF*INFO) > 125, duplicates (first occurrence was taken), consistency of reported alleles with the reference set and SE > 0 and < 10000. As different measurement instruments were used across cohorts, the meta-analysis was based on p-values and performed in the METAL software (option SCHEME SAMPLESIZE) with application of genomic control to the results of the individual cohorts. Meta-analysis results were filtered on a total sample size > 10,000. A p-value < $5E^{-8}$ was considered genome-wide significant.

Gene-based and pathway analyses. A gene-based test was performed in VEGAS,[32] with the HapMap CEU panel as a reference set (note that this excludes a large number of SNPs from the meta-analysis as cohorts imputed their data with the 1000 Genomes reference set; VEGAS is not yet updated to include information from the latter reference set). Results from the meta-analysis were included only if based on at least 90% of the total sample size (n > 15,804). All SNPs in a gene were included to obtain gene-based p-values. Genes with a p-value in the lowest quartile were then taken further to test for enrichment in biological pathways. Three different pathway analyses were run. Enrichment of pathways of biological processes and cellular localizations annotations were tested with the DAVID software (Gene Ontology, "FAT" selected subset).[33-35] In addition, an analysis was run with information from mouse knock-out phenotypes, in which we tested both specific nervous system, behavior and neurological gene sets and 30 overarching phenotypes. For each abnormal nervous system or

behavior/neurological mouse phenotype, 1-1 human orthologues of the corresponding mouse genes were considered as a pathway, following a method described by Webber et al.[36] These data were retrieved from the Mouse Genome Database (MGD),[37] Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, Maine. World Wide Web (URL: http://www.informatics.jax.org). (January, 2014). A Benjamini-Hochberg False-Discovery Rate (FDR) of 5% was used as significance threshold for the pathway analyses.[38]

GCTA. The GCTA software was used to obtain an estimate of the variance in ADHD symptoms that can be attributed to variants included in GWA studies.[39, 40] Individual level genotype data from the Netherlands Twin Register were imputed against the 1000 Genomes reference set after appropriate quality control (for more information, see [12]). Genetic variants were subsequently filtered on MAF > 0.01 and imputation quality (r^2) > 0.6 and used to estimate a genetic relationship matrix (GRM), which contains a measure of genetic similarity across all possible pairs of individuals. CBCL Attention Problem scores were used as an outcome measure, with separate analyses of mother ratings at preschool and school age, and teacher ratings at school age. For each analysis, a set of unrelated individuals was selected (genetic relatedness < 0.05). A linear mixed model was then fit to the data in which the GRM was included as a random effect and sex, age at measurement and 5 Principal Components were included as fixed effects in the model.

Results

GWA meta-analysis A summary of the QC metrics for each cohort can be found in Table 2. There was little evidence for population stratification in the results of the individual cohorts (all lambda's <= 1. 09) and the meta-analysis (lambda = 1.02).

Table 2. Results of the data cleaning for the nine cohorts included in the meta-analysis

Cohort	N	N variants uploaded	N variants cleaned	lambda
ALSPAC	5,757	31326386	8307884	1.011
Generation R	2,211	31337615	6939416	1.021
GINI / LISA	1,389	16275553	6428573	1.022
INMA	804	17405727	5593205	1.093
MOBA	665	14154076	5175847	1.018
NTR	1,605	9160231	6159203	1.027
Raine	1,338	29832393	5798626	0.986
TEDS	2,606	12223562	7134263	0.988
TRAILS	1,285	18183428	6443944	1.024

Figure 1 shows the QQ-plot of the meta-analysis filtered on results that were based on at least 10,000 individuals. Again, there was no evidence of population-stratification. The Manhattan plot in Figure 2 shows that there were no genome-wide hits. However, in the QQ-plot, there is a lift-off from the expected line for the smallest p-values that could be indicative of the polygenic nature of the trait, with many variants of small effect influencing ADHD symptoms.

Figure 1. QQ-plot of all meta-analysis results based on at least 10,000 individuals.

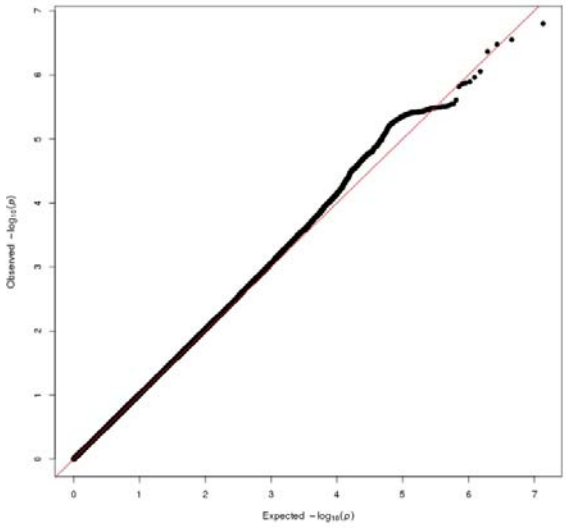
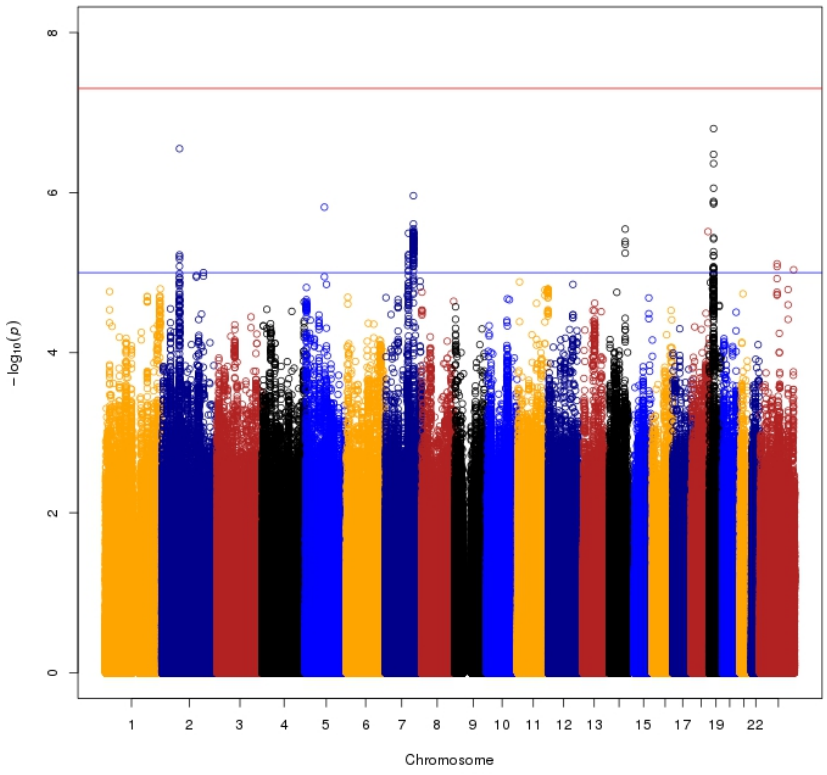


Figure 2. Manhattan plot of the meta-analysis results based on at least 10,000 individuals.



A summary of the top-signals that surpassed a threshold of suggestive association at $p < 1E^{-5}$ included in Table 3; six out of ten top variants were located in intragenic regions.

Table 3. Top variants, list of independent signals with a p-value < 1E-5 and sample size > 10,000

Genetic variant	Chr	position (b37)	Effect/other allele	Freq/freq in Refset	total N	Direction of effect	p-value	Location to nearest gene
rs56159542	19	19682971	T/C	0.21 (0.18)	17,560	- (---------)	$1.585E^{-7}$	intronic PBX4
rs2860942	2	77330790	T/G	0.21 (0.21)	17,560	- (---------)	$2.824\ E^{-7}$	intronic LRRTM4
rs144768233	7	123318554	I/R	0.47 (0.45)	17,560	+ (+++++++++)	$1.092\ E^{-6}$	downstream WASL
rs66692521	5	83303821	T/C	0.34 (0.31)	17,560	- (---------)	$1.519\ E^{-6}$	intronic EDIL3
rs79162905	14	89796072	A/G	0.11 (0.11)	17,560	- (---------)	$2.853\ E^{-6}$	intronic FOXN3
rs8083343	18	73283067	A/G	0.08 (0.09)	16,895	- (----?----)	$3.070\ E^{-6}$	intergenic
rs144888127	7	101832644	D/R	0.41 (0.39)	17,560	+ (++++++++-)	$3.214\ E^{-6}$	intronic CUX1
rs5912618	23	77764553	T/C	0.93 (0.95)	11,660	- (----????-)	$8.409\ E^{-6}$	intergenic
r s199512223	23	150863165	D/R	0.33 (0.32)	10,936	- (--?--???-)	$9.199\ E^{-6}$	upstream PRRG3
rs72891061	2	182764998	A/G	0.19 (0.21)	17,560	+ (++++-++-+)	$9.980\ E^{-6}$	intronic SSFA2

Gene-based and pathway analyses. Of the total of 17,526 genes tested in VEGAS, none surpassed the Bonferroni corrected significance threshold of $2.85E^{-6}$. The top 10 genes with the lowest gene-based p-values are listed in Table 4.

Table 4. Top 10 genes from gene-based test in VEGAS, HapMap CEU reference panel.

Gene	Chr	start position b36	stop position b36	N SNPs	p-value
LMOD2	7	123083096	123091383	82	0.00001
WASL	7	123109232	123176352	141	0.00001
ASB15	7	123036347	123065168	99	0.00002
NCOA5	20	44123032	44151987	93	0.00003
PBX4	19	19533521	19590439	62	0.00009
NCAN	19	19183781	19224061	59	0.00010
SLC12A5	20	44091244	44122196	76	0.00010
LGALS12	11	63030131	63040815	32	0.00011
EIF5A2	3	172088897	172109120	66	0.00014
RPL22L1	3	172065358	172070739	64	0.000147

In both the individual variant and gene-based association tests the top results included the genes PBX4 and WASL. Forest plots and association plots zoomed in on the regions surrounding the top variants in these genes are shown in Figure 1-4 ; association plots were created with LocusZoom.[41]

Figure 1. Forest plot of rs56159542 in PBX4. Note that the effect sizes are on standardized to allow comparison across cohorts with different measurement instruments.

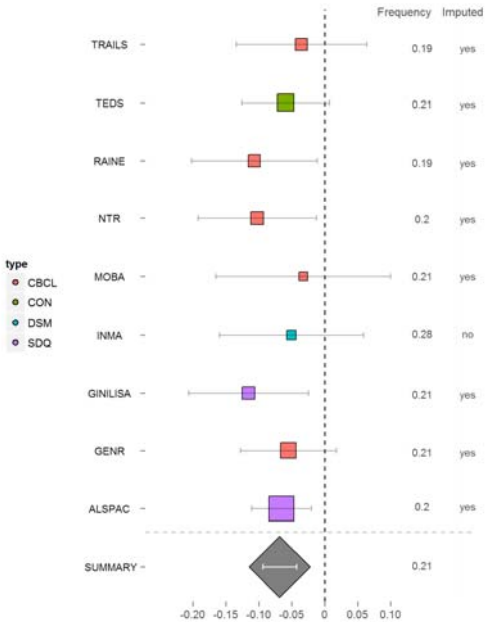


Figure 2. Forest plot of rs144768233, downstream of WASL. Note that the effect sizes are standardized to allow comparison across cohorts with different measurement instruments.

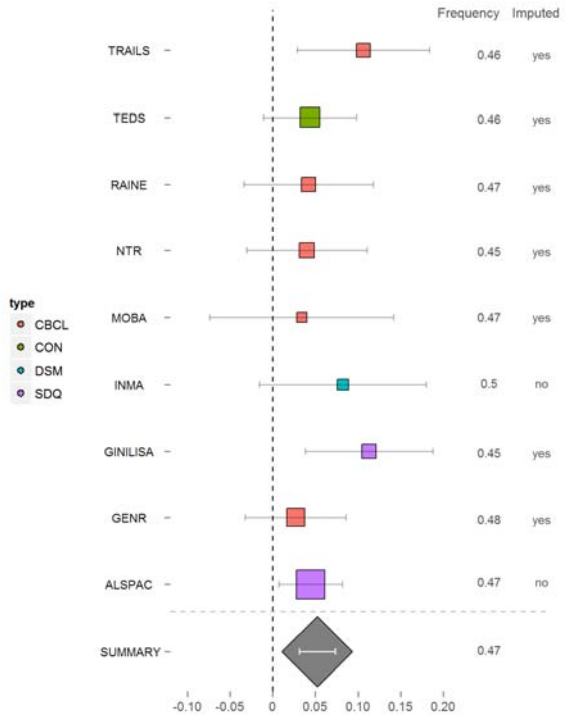


Figure 3. Association plot zoomed in on the region surrounding rs56159542 in PBX4.

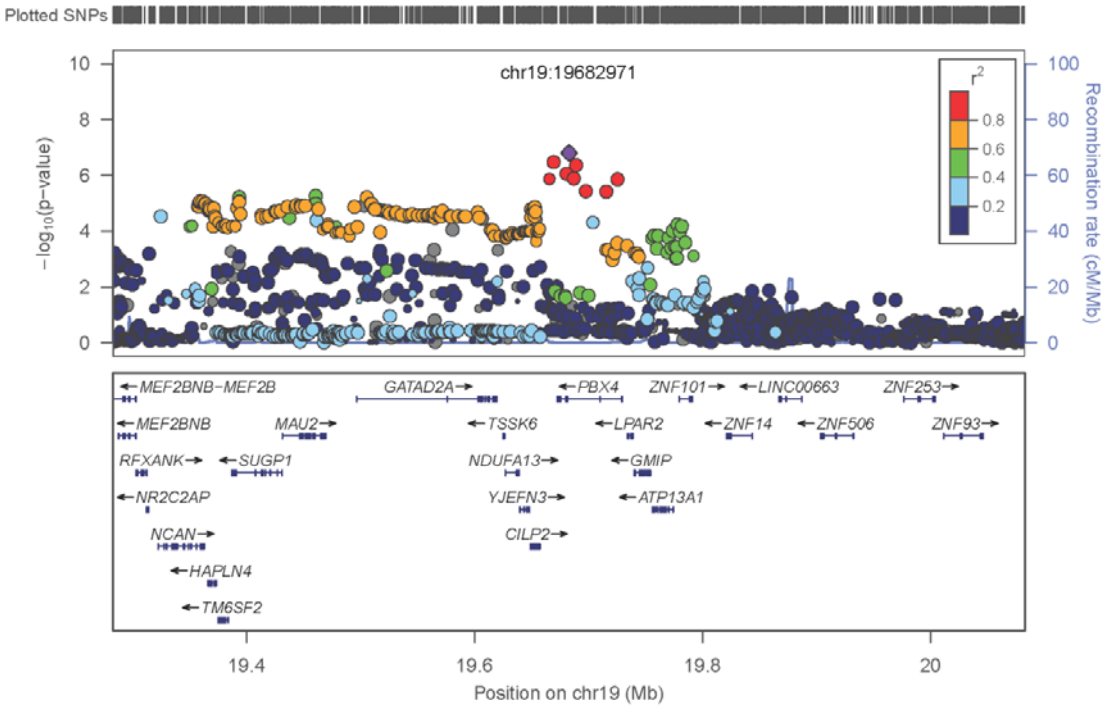
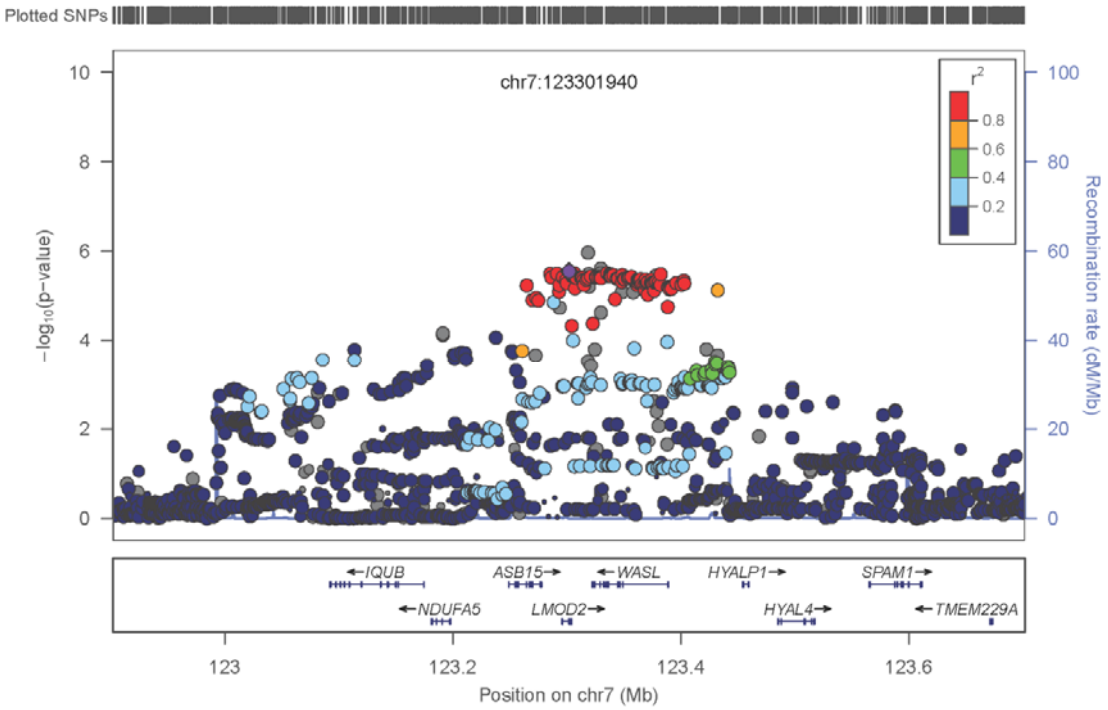


Figure 4. Association plot zoomed in on the region surrounding rs144768233, downstream of WASL. Note that information of LD is currently lacking for indels, so that the color coding is based on the variant with the second lowest p-value in this region, rs7809453.



The three pathway enrichment analyses using annotations of biological processes and cellular localizations, and mouse knockout overarching and specific neurobehavioral phenotypes did not show enrichment of a specific pathway at an FDR of 5%.

GCTA. 858-1,099 unrelated children were included in the GCTA analyses; estimates of variance in Attention Problems explained by common SNPs varied from 0.22 to 0.78 (Table 5). The variance explained in maternal school-age ratings was significantly different from zero, but it should be noted that the standard errors around the estimates are large.

Table 5. Estimates of variance in CBCL Attention Problems explained by measured genetic variants, as estimated with GCTA.

CBCL Attention Problems	N	variance explained	SE	p-value	LRT
Maternal rating preschool age	1,072	0.32	0.30	0.1	1.177
Maternal rating school age	1,099	0.78	0.30	0.004	7.036
Teacher rating school age	858	0.22	0.39	0.3	0.300

Discussion

The current study comprises the largest GWA of continuous measures of ADHD so far, and included a larger number of variants than previous studies, since all cohorts provided results for 1000 Genomes imputed datasets. In our meta-analysis of 17,560 individuals, we did not detect specific variants at genome-wide significance levels; neither did we detect significant gene-based associations or enrichment of biological pathways. Analysis with the GCTA software in a sample of ~1,000 individuals provide a first indication that common variants included in GWA studies explain variation in ADHD symptom measures in the general population.

In both the individual variant and gene-based association tests the top results included the genes PBX4 and WASL. Both genes are expressed in the brain; PBX proteins play a role in the coordination of gene expression programs during development,[42] whereas WASL is involved in cytoskeletal organization during neuronal development, including long spine formation and neurite extension.[43] While we did not find evidence for enrichment of specific biological pathways in our analyses, a previous study found enrichment of genes involved in directed neurite outgrowth when analyzing the combined results of published GWA studies on ADHD; a finding that seems to overlap with our result for WASL.[44]

The largest GWA of clinical cases of ADHD so far has been performed by the PGC. With 5,621 cases and 13,589 controls, this GWA mega-analysis had roughly the same power as the current study, but no variants reached genome-wide significance.[13] In general, GWA studies on psychiatric phenotypes have proven to require larger sample sizes than non-psychiatric phenotypes to find genome-wide results. Where early large-scale GWA studies were often not successful, recently important progress has been made in analyses of larger sample sizes; more than a hundred genome-wide significant findings have now been reported for schizophrenia when including 35,476 cases and 46,839 controls.[45] For bipolar disorder, 8 genome-wide significant hits were reported in a GWA study that included 13,741 cases and 19,762 controls.[46] If one extrapolates the success of the bipolar disorder study to ADHD, one would need 58,000 individuals with continuous measures of ADHD symptoms, or 25,000 cases and 25,000 controls to detect genetic variants at genome-wide significant levels (assuming a normal distribution of the data, following formula's for the Non Centrality Parameter (NCP) provided by Yang et al.[13]). Clearly, current sample sizes are still not approximating these numbers and substantial efforts will be required to reach them.

Population-based cohorts such as the ones collaborating in the EAGLE consortium could make an important contribution to the desired increase in power. The current study can thus be valued as an important first step towards the use of population-based cohorts in psychiatric gene-hunting studies.

A complication of the combination of results from population-based studies is the heterogeneity in measurement instruments that are typically used. Across the nine cohorts included in the current meta-analysis, there are considerable differences in both the items included in the symptom scales, the raters that provided the assessment and the ages at which the rating was collected. Phenotypic correlations across instruments, raters and time are often moderate. Yet, there is evidence from behavior genetic studies that genetic factors influence partly instrument, rater and age-specific behavior, but also a construct that is common across these measures.[47-51] This was confirmed by a polygenic risk score analysis that indicated a genetic overlap between the aggregate genetic signals in the latest PGC ADHD GWA mega-analysis and both parent and teacher ratings, and preschool and school-age ratings of Attention Problems in the Netherlands Twin Register (NTR).[12] It seems therefore plausible that the effects of increased phenotypic heterogeneity will be outweighed by the increase in sample size and hence statistical power. In addition, statistical methods like Item Response Theory (IRT) could be used to synchronize cohorts with different measurement instruments in a sophisticated manner; this approach has for example been successfully applied in a GWA meta-analysis of personality measures.[52,53]

In this context it is noteworthy that issues of heterogeneity are not unique to population-based cohorts; it seems likely that the small effect sizes observed in current psychiatric GWA studies are related to a larger genetic and phenotypic heterogeneity, which in turn reflect the fact that the diagnostic system is based on expert opinions and descriptions of symptoms, and may insufficiently reflect underlying biological categories. Refinement of the diagnostic system could result from studies that focus on observable behavior and neurobiological measures, such as imaging techniques or biomarkers; research that is currently explicitly stimulated by the Research Domain Criteria Project (RDoC) of the National Institute of Mental Health (NIMH).[54] If successful, GWA studies of these measures could likely increase statistical power, as they would benefit both from the full information of continuous measures and the closer reflection of biological entities.

Although current sample sizes are limited in their ability to detect individual genetic variants, they have been used successfully to assess aggregate genetic risk for ADHD with polygenic risk score analyses, and to estimate the contribution of common SNPs to overall variation in the liability to ADHD.[12, 55-57] Our GCTA analysis performed in one sample from the meta-analysis provides a first indication that common SNPs explain variance in Attention Problems. The precision of the estimate is currently still limited by the relatively small sample size. Unfortunately it is not straightforward to combine individual level genetic data from multiple cohorts; but efforts are currently made to combine the data of two cohorts from the EAGLE consortium to allow estimates of much larger precision. Two previous studies have used the GCTA software to estimate the amount of variance in ADHD liability or symptoms that can be attributed to variants included in GWA studies. Whereas the earlier mentioned study by the PGC estimated that 28% of the liability to ADHD could be explained by common SNPs, another study on self, parent and teacher ratings of ADHD symptoms by TEDS found non-significant estimates close to zero.[7, 58] As there is no obvious explanation for these differences, future studies such as the one currently performed within cohorts of the EAGLE consortium can hopefully help to clarify this issue.

References

1. Polanczyk G, de Lima MS, Horta BL, Biederman J, Rohde LA. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. Am J Psychiatry. 2007;164:942-948.

2. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Arlington, VA: American Psychiatric Publishing; 2013.

3. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed, text revision. Washington, DC: American Psychiatric Association; 2000.

4. Faraone SV, Perlis RH, Doyle AE, et al. Molecular genetics of attention-deficit/hyperactivity disorder. Biol Psychiatry. 2005;57:1313-1323.

5. Faraone SV, Mick E. Molecular genetics of attention deficit hyperactivity disorder. Psychiatr Clin North Am. 2010;33:159-180.

6. Holmans P. Pathway analysis of large multicentre ADHD genome-wide datasets. Abstract of oral presentation 21st World Congress of Psychiatric Genetics; October 17-21, 2013; Boston, USA.

7. Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee SH, Ripke S, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet. 2013;45:984-994.

8. van der Sluis S, Posthuma D, Nivard MG, Verhage M, Dolan CV. Power in GWAS: lifting the curse of the clinical cut-off. Mol Psychiatry. 2013;18:2-3.

9. Lubke GH, Hudziak JJ, Derks EM, van Bijsterveldt TC, Boomsma DI. Maternal ratings of attention problems in ADHD: evidence for the existence of a continuum. J Am Acad Child Adolesc Psychiatry. 2009;48:1085-1093.

10. Levy F, Hay DA, McStephen M, Wood C, Waldman I. Attention-deficit hyperactivity disorder: a category or a continuum? Genetic analysis of a large-scale twin study. J Am Acad Child Adolesc Psychiatry. 1997;36:737-744.

11. Larsson H, Anckarsater H, Rastam M, Chang Z, Lichtenstein P. Childhood attention-deficit hyperactivity disorder as an extreme of a continuous trait: a quantitative genetic study of 8,500 twin pairs. J Child Psychol Psychiatry. 2012;53:73-80.

12. Groen-Blokhuis MM, Middeldorp CM, Abdellaoui A, et al. Attention Deficit Hyperactivity Disorder polygenic risk scores predict Attention Problems in a

population-based sample of children. J Am Acad Child Adolesc Psychiatry. Under review.

13. Yang J, Wray NR, Visscher PM. Comparing apples and oranges: equating the power of case-control and quantitative trait association studies. Genet Epidemiol. 2010;34:254-257.

14. Benke KS, Nivard MG, Velders FP, et al. A genome-wide association meta-analysis of preschool internalizing problems J Am Acad Child Adolesc Psychiatry. 2013, accepted for publication.

15. Paternoster L, Standl M, Chen CM, et al. Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. Nat Genet. 2012;44:187-192.

16. Taal HR, St Pourcain B, Thiering E, et al. Common variants at 12q15 and 12q24 are associated with infant head circumference. Nat Genet. 2012;44:532-538.

17. Achenbach TM, Rescorla LA. Manual for the ASEBA Preschool forms & Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families; 2000.

18. Achenbach TM, Rescorla LA. Manual for the ASEBA School-Age Forms & Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families; 2001.

19. Conners CK. Conners' Rating Scales - Revised. New York, NY: Multi-Health Systems; 2001.

20. Goodman R. The Strengths and Difficulties Questionnaire: a research note. J Child Psychol Psychiatry. 1997;38:581-586.

21. Golding J, Pembrey M, Jones R, Team AS. ALSPAC--the Avon Longitudinal Study of Parents and Children. I. Study methodology. Paediatr Perinat Epidemiol. 2001;15:74-87.

22. Jaddoe VW, van Duijn CM, van der Heijden AJ, et al. The Generation R Study: design and cohort update 2010. Eur J Epidemiol. 2010;25:823-841.

23. Rzehak P, Sausenthaler S, Koletzko S, et al. Period-specific growth, overweight and modification by breastfeeding in the GINI and LISA birth cohorts up to age 6 years. Eur J Epidemiol. 2009;24:449-467.

24. Guxens M, Ballester F, Espada M, et al. Cohort Profile: the INMA--INfancia y Medio Ambiente--(Environment and Childhood) Project. Int J Epidemiol. 2012;41:930-940.

25. Magnus P, Irgens LM, Haug K, et al. Cohort profile: the Norwegian Mother and Child Cohort Study (MoBa). Int J Epidemiol. 2006;35:1146-1150.

26. van Beijsterveldt CE, Groen-Blokhuis M, Hottenga JJ, et al. The Young Netherlands Twin Register (YNTR): longitudinal twin and family studies in over 70,000 children. Twin Res Hum Genet. 2013;16:252-267.

27. Evans S, Newnham J, MacDonald W, Hall C. Characterisation of the possible effect on birthweight following frequent prenatal ultrasound examinations. Early Hum Dev. 1996;45:203-214.

28. Newnham JP, Evans SF, Michael CA, Stanley FJ, Landau LI. Effects of frequent ultrasound during pregnancy: a randomised controlled trial. Lancet. 1993;342:887-891.

29. Williams LA, Evans SF, Newnham JP. Prospective cohort study of factors influencing the relative weights of the placenta and the newborn infant. BMJ. 1997;314:1864-1868.

30. Haworth CM, Davis OS, Plomin R. Twins Early Development Study (TEDS): a genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. Twin Res Hum Genet. 2013;16:117-125.

31. Huisman M, Oldehinkel AJ, de Winter A, et al. Cohort profile: the Dutch 'TRacking Adolescents' Individual Lives' Survey'; TRAILS. Int J Epidemiol. 2008;37:1227-1235.

32. Liu JZ, McRae AF, Nyholt DR, et al. A versatile gene-based test for genome-wide association studies. Am J Hum Genet. 2010;87:139-145.

33. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25-29.

34. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4:44-57.

35. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37:1-13.

36. Webber C, Hehir-Kwa JY, Nguyen DQ, de Vries BB, Veltman JA, Ponting CP. Forging links between human mental retardation-associated CNVs and mouse gene knockout models. PLoS Genet. 2009;5:e1000531.

37. Eppig JT, Blake JA, Bult CJ, Kadin JA, Richardson JE, Mouse Genome Database G. The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse. Nucleic Acids Res. 2012;40:D881-886.

38. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B (Methodological). 1995;57:289–300.

39. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76-82.

40. Yang J, Manolio TA, Pasquale LR, et al. Genome partitioning of genetic variation for complex traits using common SNPs. Nat Genet. 2011;43:519-525.

41. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010;26:2336-2337.

42. Laurent A, Bihan R, Omilli F, Deschamps S, Pellerin I. PBX proteins: much more than Hox cofactors. Int J Dev Biol. 2008;52:9-20.

43. Takenawa T, Suetsugu S. The WASP-WAVE protein network: connecting the membrane to the cytoskeleton. Nat Rev Mol Cell Biol. 2007;8:37-48.

44. Poelmans G, Pauls DL, Buitelaar JK, Franke B. Integrated genome-wide association study findings: indetification of a neurodevelopmental network for Attention Deficit Hyperactivity Disorder. Am J Psychiatry. 2011;168:365-377.

45. Ripke S, Psychiatric Genomics Consortium Schizophrenia Working Group. Psychiatric Genomics Consortium quadruples schizophrenia GWAS sample-size to 35,000 cases and 47,000 controls. Abstract of oral presentation 21st World Congress of Psychiatric Genetics; October 17-21, 2013; Boston, USA.

46. Charney AW, Ruderfer D, Moran J, et al. Bipolar disorder GWAS of 13,741 cases and 19,762 controls identifies eight genome-wide significant hits and implicates genes targeted by FMRP protein. Abstract of oral presentation 21st World Congress of Psychiatric Genetics; October 17-21, 2013; Boston, USA.

47. Derks EM, Hudziak JJ, Beijsterveldt CE, Dolan CV, Boomsma DI. Genetic analyses of maternal and teacher ratings on attention problems in 7-year-old Dutch twins. Behav Genet. 2006;36:833-844.

48. Derks EM, Hudziak JJ, Dolan CV, Beijsterveldt TC, Verhulst FC, Boomsma DI. Genetic and environmental influences on the relation between attention problems and attention deficit hyperactivity disorder. Behav Genet. 2008;38:11-23.

49. Kan K, Dolan CV, Nivard MG, et al. Genetic and environmental stability in attention problems across the lifespan: evidence from the Netherlands twin register. J Am Acad Child Adolesc Psychiatry. 2013;52:12-25.

50. Kuntsi J, Rijsdijk F, Ronald A, Asherson P, Plomin R. Genetic influences on the stability of attention-deficit/hyperactivity disorder symptoms from early to middle childhood. Biol Psychiatry. 2005;57:647-654.

51. Thapar A, Harrington R, Ross K, McGuffin P. Does the definition of ADHD affect heritability? J Am Acad Child Adolesc Psychiatry. 2000;39:1528-1536.

52. De Moor MHM, Van den Berg SM, Genetics of Personality Consortium, Boomsma DI. The influence of common genetic variants on Neuroticism and their overlap with Major Depression: Evidence from the Genetics of Personality Consortium. In preparation.

53. Van den Berg SM, De Moor MHM, Genetics of Personality Consortium, Boomsma DI. Harmonization of Neuroticism and Extraversion phenotypes across inventories and cohorts in the Genetics of Personality Consortium: an application of Item Response Theory. Behavior Genetics. Under review.

54. Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. BMC Med. 2013;11:126.

55. Cross-Disorder Group of the Psychiatric Genomics C, Genetic Risk Outcome of Psychosis C. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet. 2013;381:1371-1379.

56. Hamshere ML, Langley K, Martin J, et al. High loading of polygenic risk for ADHD in children with comorbid aggression. Am J Psychiatry. 2013;170:909-916.

57. Hamshere ML, Stergiakouli E, Langley K, et al. Shared polygenic contribution between childhood attention-deficit hyperactivity disorder and adult schizophrenia. Br J Psychiatry.203:107-111.

58. Trzaskowski M, Davis OS, DeFries JC, Yang J, Visscher PM, Plomin R. DNA evidence for strong genome-wide pleiotropy of cognitive and learning abilities. Behav Genet. 2013;43:267-273.

Supplementary Materials

Supplementary Table 1. Description of methods used for imputation and analysis in each cohort included in the GWA meta-analysis.

Cohort	Genotyping platform	Pre-imputation variant filters				Pre-imputation sample filters					Imputation software	Post-imputation filters	Association software
		call rate	MAF	HWE	other filers	call rate	heterozygosity	etnicity	gender	other filers			
ALSPAC	Illumina HumanHap550 quad-chip	0.95	0.01	5E-7		0.97	yes	yes	yes	> 10% identity by descent, insufficient sample replication	Phasing with Mach, Imputation with Minimac	None	Mach2QTL V112
Generation R	Illumina Human 610 and 660 Quad Array	0.95	0.001	1E-7		0.975	yes	yes	yes	familial relationship mismatch	Mach	None	Plink 1.07
GINI / LISA	Affymetrix 5.0 and Affymetrix 6.0	0.95	0.01	1E-5		0.95	> 4 SD		yes	similarity QC based on MDS	Impute v2.3.0	SNPTEST NA for BETA, SE and P_VAL	SNPTEST v2.4.1
INMA	Illumina Human Omni1	0.95	0.01	1E-6		0.98		no	yes	LRR SD > 0.3, duplicates and relatedness checked	Impute v.2	None	SNPtest v.2
MOBA	Illumina Human 660W Quad Array	0.97	0.01	1E-6	Mitochodrial SNPs; chrY&PAR SNPs; SNPs that could not be updated to hg37; non-"rs" SNPs	0.96	> 4 SD	yes	yes	Cryptic Relatedness	SHAPEIT (v2.r644), Impute (version 2.3.0)	SNPTEST NA for BETA and P_VAL	SNPTEST v2.5-beta4
NTR	Affymetrix 6.0	0.95	0.01	1E-5	double typed error rate > 0.02; Mendel error rate > 0.02;	0.90	F > 0.10 or F < -	no	yes	IBS/IBD discrepencies; Mendel error	Minimac and Mach	Plink NA for BETA, SE and	Plink 1.07

					allele frequency frequency difference with reference set > 0.20; C/G and A/T SNPs with MAF > 0.35		0.10			rate > 0.02		P_VAL	
Raine	Illumina Human 660W Quad Array	0.95	0.01	5.7E-7	C/G and A/T SNPs removed	0.97	F > 0.1875; heterozygosity > 0.30	no	yes		Mach		
TEDS	Affymetrix 6.0	0.80	0.01	1E-20	SNP plate association p-value < 10^-6, SNPTEST info > 0.975; visual inspection of hybridization intensity plots	yes	yes	yes	yes	unusual hybridization intensity, relatedness (IBD < 5%), regenotyping low concordance	Impute v2	None	Plink 1.07
TRAILS	Illumina Cyto SNP12 v2	0.95	0.01	1E-3	chr X SNPs > 1% heterozygous in men	0.95	> 4 SD	yes	yes	duplicates	Impute v2	Callrate 10%, duplicates	SNPtest 2.4.1

Supplementary Table 2. Item content of ADHD symptom scales included in the GWA meta-analysis.

CBCL 1.5-5: Attention Problems scale	• Can't concentrate, can't pay attention for long • Can't sit still, restless, or hyperactive • Poorly coordinated or clumsy • Quickly shifts from one activity to another • Wanders away
CBCL 6-18: Attention Problems scale	• Acts too young for his/her age • Fails to finish things he/she starts • Can't concentrate, can't pay attention for long • Can't sit still, restless, or hyperactive Confused or seems to be in a fog • Daydreams or gets lost in his/her thoughts • Impulsive or acts without thinking • Poor school work • Inattentive or easily distracted • Stares blankly
SDQ: Hyperactivity-inattention scale	• Restless, overactive, cannot stay still for long • Constantly fidgeting or squirming • Easily distracted, concentration wanders • Thinks things out before acting • Sees tasks through to the end, good attention span
Conners' Rating Scales-Revised: Long Form Other DSM-IV based rating scales (items may be phrased slightly differently across scales)	• Often fails to give close attention to details or makes careless mistakes in schoolwork, work, or other activities • Often has difficulty sustaining attention in tasks or play activities • Often does not seem to listen when spoken to directly • Often does not follow through on instructions and fails to finish schoolwork, chores, or duties in the workplace (not due to oppositional behaviour or failure of comprehension) • Often has difficulty organizing tasks and activities • Often avoids, dislikes, or is reluctant to engage in tasks that require sustained mental effort (such as schoolwork or homework) • Often loses things necessary for tasks or activities at school or at home (e.g. toys, school

	assignments, pencils, books or tools) • Is often easily distracted by extraneous stimuli • Is often forgetful in daily activities • Often fidgets with hands or feet or squirms in seat • Often leaves seat in classroom or in other situations in which remaining seated is expected • Often runs about or climbs excessively in situations in which it is inappropriate • Often has difficulty playing or engaging in leisure activities quietly • Is often 'on the go' or often acts as if 'driven by a motor' • Often talks excessively • Often has difficulty awaiting turn • Often blurts out answers to questions before they have been completed • Often interrupts or intrudes on others, e.g. butts into other children's games

Chapter 7

Developmental Trajectories in Twins (DETECT): an overview of genotyping, data cleaning, imputation and analyses

Groen-Blokhuis MM, Willemsen AHM, Ehli EA, Abdellaoui A, Hottenga JJ, Scheet PA, Middeldorp CM, Boomsma DI

During my PhD thesis, a large genotyping project was undertaken in which samples from the NTR were genotyped in collaboration with three USA based institutions: the University of Vermont, the Avera Institute for Human Genetics and the University of Houston. The data that were generated in this project were subsequently used for genetic analyses; the results of the analyses on attention problems are described in chapter four, five and six of this thesis. This chapter aims to describe the results of genetic analyses on other phenotypes that were performed for international consortia, as well as the data generation, cleaning and imputation steps that were performed in order to enable these analyses.

Content

Abbreviations (in alphabetical order)

ANTR	Adult Netherlands Twin Register
ARRA	American Recovery and Reinvestment Act
CEU	Utah Residents with Northern and Western European ancestry
CNV	Copy Number Variant
CQC	Contrast Quality Control
dbGaP	Database of Genotypes and Phenotypes
DZ	Dizygotic
TdT	Terminal Deoxynucleotidyl Transferase
DETECT	Developmental Trajectories in Twins
EAGLE	EArly Genetics and Lifecourse Epidemiology

EGG	Early Growth Genetics
GO	Grand Opportunity
GoNL	Genome of the Netherlands
GWA	Genome Wide Association
GWAS	Genome Wide Association Study
HWE	Hardy-Weinberg equilibrium
IBD	Identity by Descent
MAF	Minor Allele Frequency
ME	Mendelian Errors
MZ	Monozygotic
NESDA	Netherlands Study of Depression and Anxiety
NIMH	National Institute of Mental Health
NTR	Netherlands Twin Register
PCA	Principal Component Analysis
PC	Principal Component
PCR	Polymerase Chain Reaction
PI	Principal Investigator
QC	Quality Control
RPM	Revolutions Per Minute
SAPE	Steptavidin Phycoerythrin
SNP	Single Nucleotide Polymorphism
USA	United States of America
VNTR	Variable Number of Tandem Repeats
YNTR	Young Netherlands Twin Register

I. Genome Wide Association analyses

Context. All GWA analyses using DETECT data were performed within two closely collaborating European consortia that perform GWA meta-analysis projects in large population based cohorts focusing on childhood phenotypes. The EGG consortium focuses on traits related to early growth and the EAGLE consortium focuses on a broad range of childhood phenotypes such as blood pressure, eczema, and behavioral phenotypes. More information on the genotyping, cleaning and imputation procedures performed on the DETECT set is provided in paragraphs II-VIII.

1. Genome-wide association study on atopic dermatitis – HapMap imputation

The first imputation of the DETECT set was performed in order to participate in a GWA meta-analysis on atopic dermatitis in childhood within the EAGLE consortium. This GWA meta-analysis project had finished the discovery phase and was seeking replication of the top 10 SNPs that came out of the discovery phase. Data on eczema were available from a YNTR survey that is sent out to the parents of twins at age 5. Parents were asked to indicate for each child separately whether a doctor had ever diagnosed the child with eczema. A similar question concerned doctor diagnosed baby eczema. Children were considered cases if their parents answered yes to any of the two questions and controls if they answered no to both questions. One twin was selected from each family. If both twins were cases or controls, one individual was picked at random, otherwise the case was selected. A total of 123 cases and 306 controls were included in the study. A logistic regression was performed in Plink in which eczema status was regressed on SNP dosage scores in an additive model with sex as a covariate. SNP dosage scores represent the estimated expected allele count for the effect allele on a scale from zero to two and are used to take into account the uncertainty of the imputation in the model. The results for the 10 replication SNPs are listed in Table 1. In the final meta-analysis in which the discovery and replication studies were combined (total n = 51,423), three SNPs reached genome-wide significance. Two SNPs (rs479844 and rs2164983) were near genes that have been implicated in epidermal proliferation and differentiation and one SNP (rs2897442) was located within the cytokine cluster at 5q31.1.[1]

Table 1. Results of the logistic regression of eczema status on the 10 replication SNPs for the NTR.

SNP	Effect allele	Effect allele frequency	Beta	SE	P-value
rs7000782	A	0.3747	-0.14156	0.1716	0.4095
rs1327914	C	0.1799	0.074922	0.1917	0.6959
rs2164983	A	0.1461	0.5652	0.2173	0.009296
rs479844	A	0.4558	-0.06646	0.1537	0.6657
rs10994675	A	0.4371	-0.05077	0.1492	0.7335
rs4821544	C	0.2716	0.15965	0.1766	0.3659
rs10983837	A	0.0333	0.323387	0.6118	0.5971
rs2897442	C	0.2746	0.194085	0.1704	0.2546
rs3853601	C	0.8583	-0.71805	0.2327	0.002031
rs4520482	A	0.4732	-0.00844	0.1507	0.9551

2. Genome-wide association study on atopic dermatitis – 1000 Genomes imputation

After the initial EAGLE atopic dermatitis GWA meta-analysis was published,[2] a follow-up project was started in which cohorts were requested to run the analyses on SNP data imputed against the 1000 Genomes haplotypes reference panel for all ancestries. For the NTR, genotype data as available in MRG5 were used and the same phenotype definition as in the previous replication analysis applied. For MZ twin pairs, one individual was selected at random, for DZ twin pairs, both twins were included in the analysis while correcting for familial clustering with the --family option in Plink. After exclusion of individuals with non-Dutch/non-European ancestry, 1,466 individuals were available for analysis. A logistic regression was performed in Plink in which eczema status was regressed on genome-wide SNP data, sex, the 3 Dutch PCs, one chip covariate and the buccal PC. To test whether the results showed any systematic inflation, the inflation factor lambda (the observed median chi-square divided by the expected median chi-square) was calculated and a QQ-plot was drawn. The lambda of the cleaned results was 1.017. The QQ and Manhattan plot of all SNPs cleaned on MAF > sqrt(5/1466) and Plink INFO > 0.4 are shown in Figure 1 and 2. The discovery phase of this project has now been closed and the analysis team is working on the meta-analysis.

Figure 1. QQ-plot of the NTR GWAS on eczema status with SNPs cleaned on MAF > sqrt(5/1466) and Plink INFO > 0.4.

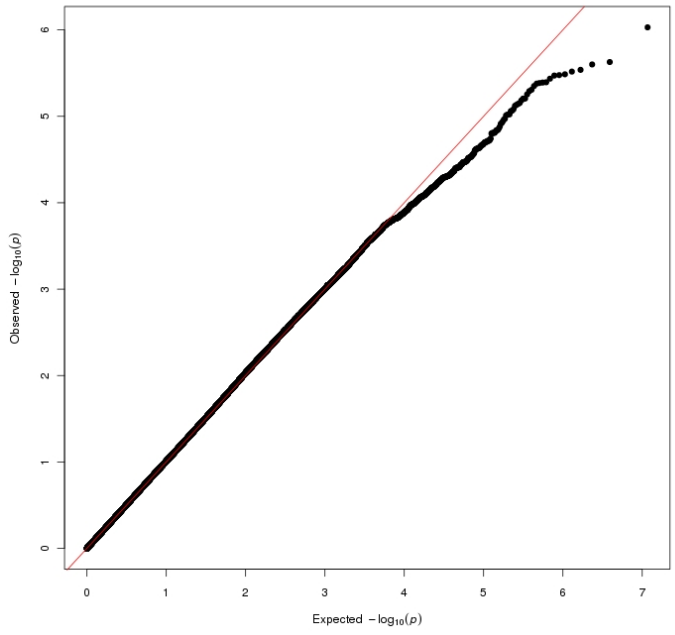
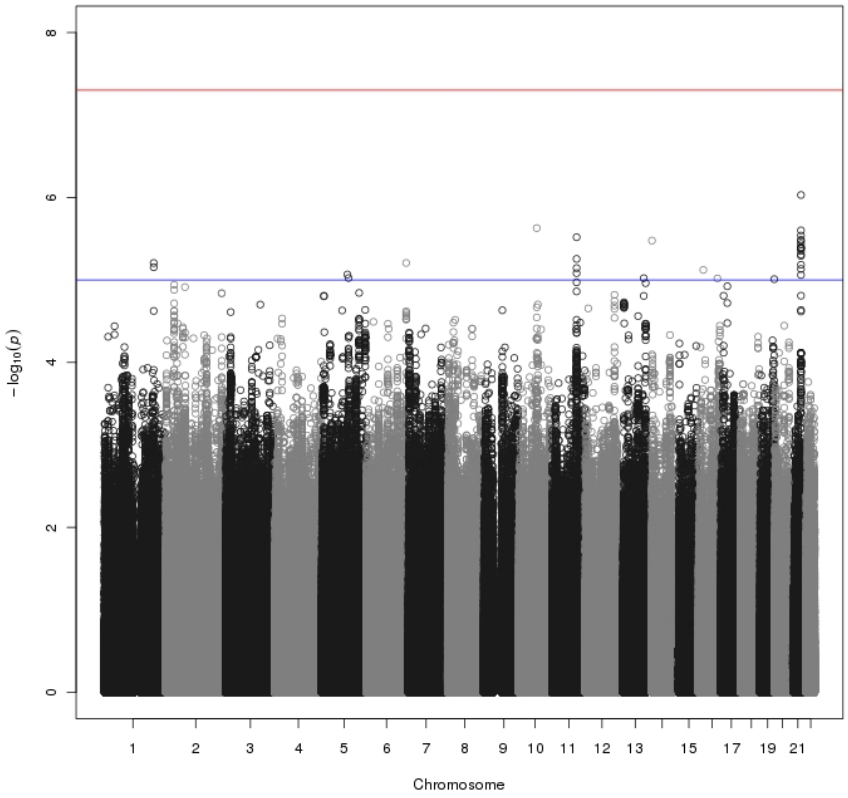


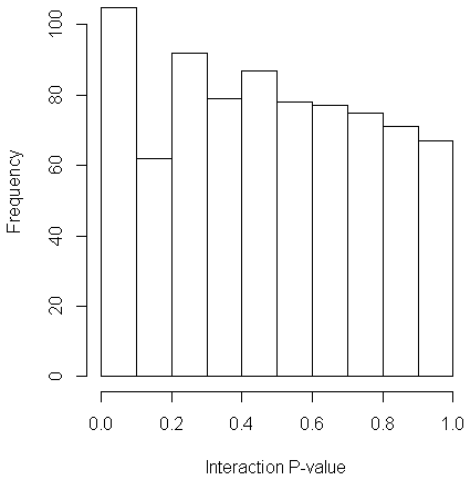
Figure 2. Manhattan plot of the NTR GWAS on eczema status with SNPs cleaned on MAF > sqrt(5/1466) and Plink INFO > 0.4.



3. Epistatic interaction study of known atopy and atopic dermatitis GWAS hits on atopic dermatitis

As a side project to the 1000 Genomes atopic dermatitis GWAS, an effort was made to analyze epistatic interactions among SNPs that were previously reported as top hits in GWAS on eczema and other atopic diseases. Mutations in the Filaggrin gene were of special interest in the study, but were unfortunately not available for the NTR. Of the 42 remaining SNPs selected by the consortium, 41 were imputed in MRG5 (one SNP was only available for cohorts genotyped on the Illumina platform). SNPs were extracted from the MRG5 set and dosage scores were calculated for those SNPs such that the dosage score reflected the number of risk alleles (Linux and Pearl script provided by Jouke-Jan Hottenga and Andrew R. Wood). The phenotype data were the same as for the GWA analysis. SNP-by-SNP interactions were then tested in Stata; a logistic regression was performed in which eczema status was regressed on SNP1, SNP2, SNP1-by-SNP2, sex, the 3 Dutch PCs, one chip covariate and the buccal PC, while correcting for familial clustering with the "robust cluster" option. For two SNPs with a low minor allele frequency (rs987870 and rs6932730), some interaction analyses could not be performed due to the empty cells that occurred. In, total 793 interactions were tested. As can be seen in Figure 3, a relative excess of small p-values were observed for the interaction analyses in the NTR dataset. The analysis team has finished the data cleaning and is performing a meta-analysis of all results.

Figure 3. Histogram of the p-values of the SNP-by-SNP interactions tested in the NTR sample.



4. Genome-wide association study on height at age one

Within the EGG consortium various GWAS have been initiated that focus on growth in the early years of life. In one specific project, a GWA meta-analysis was planned for height at age one, two, three and four. Initially, analyses for the NTR have been performed for all these ages on the imputed set of February 2011. When the imputed set of September 2011 became available, the consortium had decided to focus solely on height at age 1 and these analyses are reported here. Data on height at age one were available from maternal reports and laboratory measures. Maternal reports were collected in the YNTR survey that is sent out at age two. Mothers were asked to copy the measurements on height and weight that have been performed at the Dutch Community Health Services at a regular basis during the first two years of life. Laboratory measures were available from smaller projects focusing on cognition and physical development.[3] Improbable data points were checked for data entry errors and corrected where possible. For each individual, the measurement between age 6-18 months closest to age 12 months was selected. When both were available, laboratory measures were preferred over maternal reports. Sex- and age standardized scores were calculated in the software package Growth analyser 3 (2004) using the 1997 Dutch reference growth charts for the general population. After exclusion of individuals with non-Dutch / non-European ancestry, 1,674 children were available for analysis. Analyses were performed in Plink; standardized height scores were regressed on genome-wide SNP data and the PC that correlated with a north-south gradient in the Netherlands, as the latter was significantly correlated with the phenotype. The lambda of the cleaned results (MAF > sqrt(5/1674) and Plink INFO > 0.4) was 1.004. Manhattan and QQ plots are provided in Figure 4 and 5. This project has closed the discovery phase and the meta-analysis is currently in progress.

Figure 4. QQ-plot of the NTR GWAS on height at age 1, SNPs cleaned on MAF > sqrt(5/1674) and Plink INFO > 0.4.

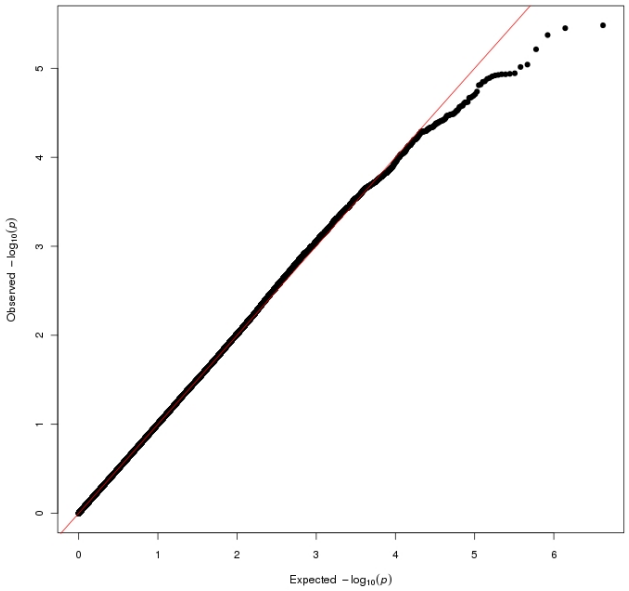
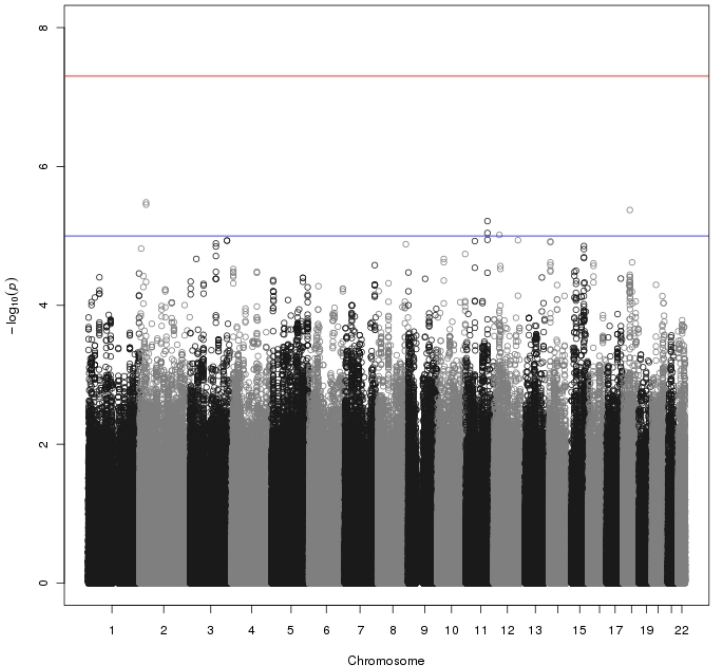


Figure 5. Manhattan plot of the NTR GWAS on height at age 1, SNPs cleaned on MAF > sqrt(5/1674) and Plink INFO > 0.4.



5. Genome-wide association study on BMI in childhood

Within EGG, another GWA meta-analysis project focused on BMI in childhood. Three age-bins were defined: BMI at 4 years (24-60 months), BMI at 9 years (60-120 months) and an overall analysis in which the measurement point between 24 and 120 months closest to 120 months was selected. In the NTR, data on BMI were available from the surveys sent out at age 2, 3, 5, 7 and 10. At age 2 and 3, parents were asked to fill out all the dates and measures of height and weight as measured by the Dutch Community Health Services at regular intervals. At age 5, parents were asked to fill out the height and weight of the twins with the dates of measurement since the 3th birthday. At age 7 and 10, parents were asked to fill out the current height and weight of the twins with the date of assessment. In addition, laboratory measures were available for a group of individuals that participated in smaller subprojects.[3] Improbable data points were checked for data entry errors and corrected or excluded where necessary. Sex- and age adjusted standardized BMI scores were calculated with the 1990 British growth reference charts using the program LMSGrowth (available from http://www.healthforallchildren.co.uk). Individuals with non-Dutch / non-European ancestry were excluded from the analyses, which resulted in a sample size of respectively 1,664, 1,480 and 1,810 individuals for the analysis of BMI at age 4, age 9 and the overall analysis. Analyses were performed in Plink in a linear regression framework; age and sex adjusted standard deviation scores were regressed on genome-wide SNP data and the first 2 principal components, as this was requested in the analysis protocol. Standard errors were adjusted for family clustering using the --family option in Plink. The results of the NTR analyses are shown in Figure 6-11, the lambda's of the cleaned results of the 3 analyses were 1.015, 1.017 and 1.005 (cleaned on MAF > sqrt(5/total sample size) and INFO > 0.4). A preliminary meta-analysis was reported in December 2013; additional cohorts have been contacted for replication of the top signals. Simultaneously, cohorts from the discovery phase will be asked to rerun the analyses on datasets imputed against the 1000 Genomes reference set.

Figure 6. QQ-plot of the NTR GWAS on BMI at 4 years (24-60 months), SNPs cleaned on MAF > sqrt(5/1664) and Plink INFO > 0.4.

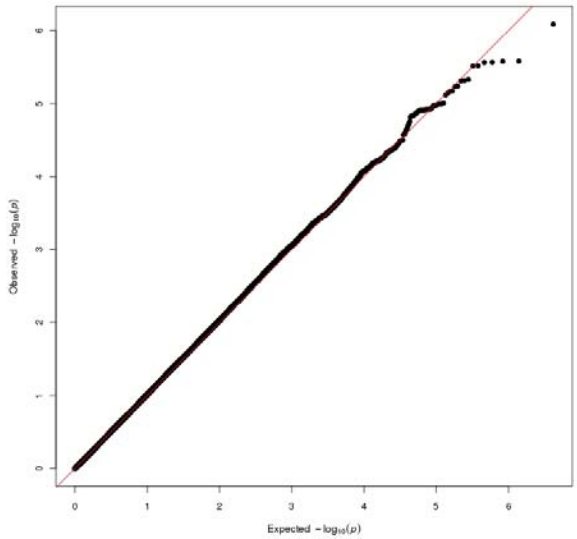


Figure 7. Manhattan plot of the NTR GWAS on BMI at 4 years (24-60 months), SNPs cleaned on MAF > sqrt(5/1664) and Plink INFO > 0.4.

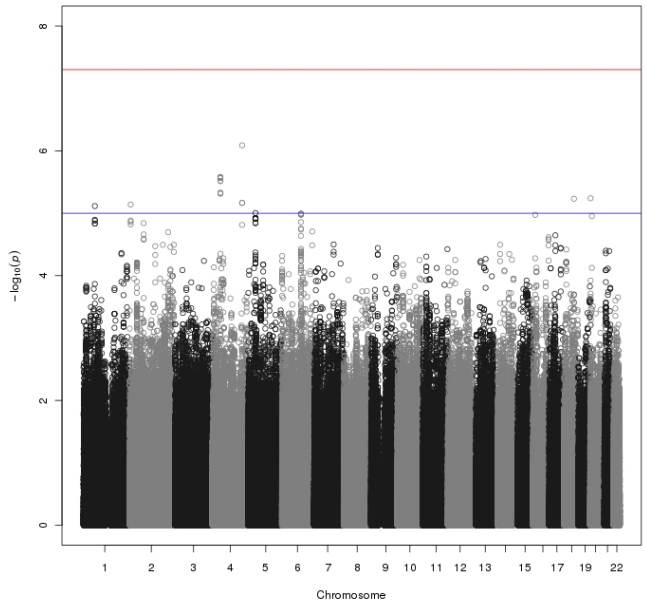


Figure 8. QQ-plot of the NTR GWAS on BMI at 9 years (60-120 months) with SNPs cleaned on MAF > sqrt(5/1480) and Plink INFO > 0.4.

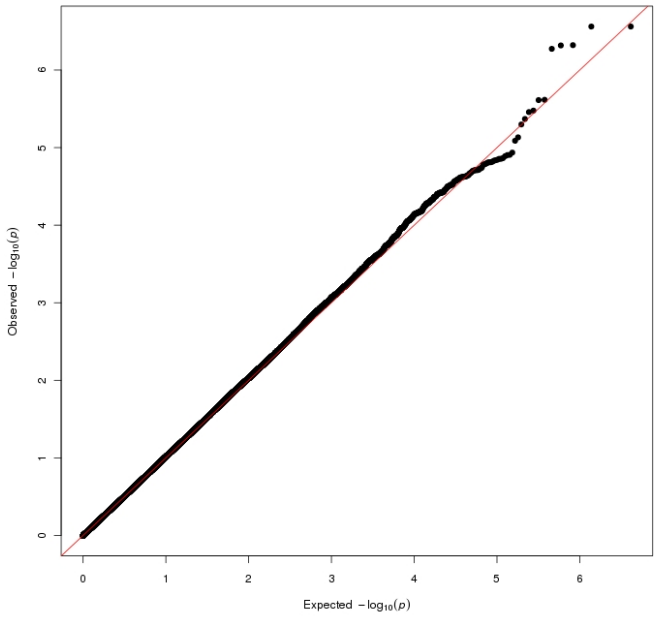


Figure 9. Manhattan plot of the NTR GWAS on BMI at 9 years (60-120 months) with SNPs cleaned on MAF > sqrt(5/1480) and Plink INFO > 0.4.

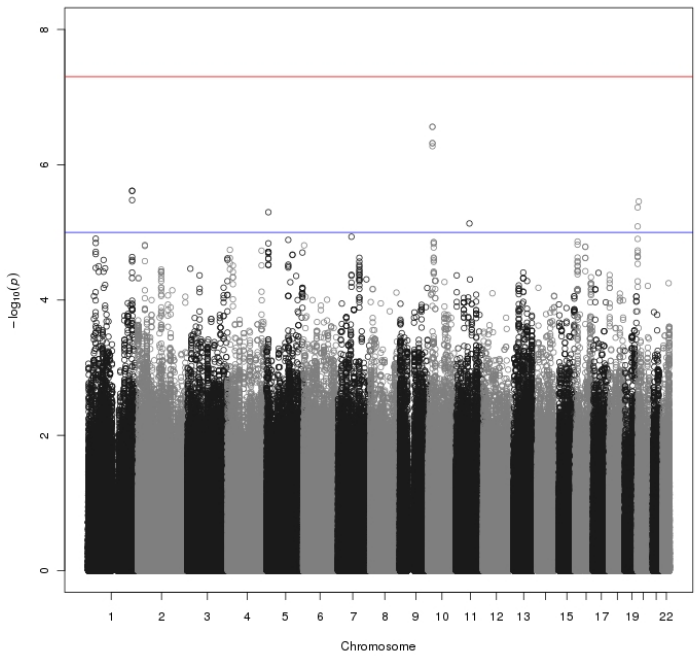


Figure 10. QQ-plot of the NTR GWAS on BMI at the latest age between 24-120 months with SNPs cleaned on MAF > sqrt(5/1810) and Plink INFO > 0.4.

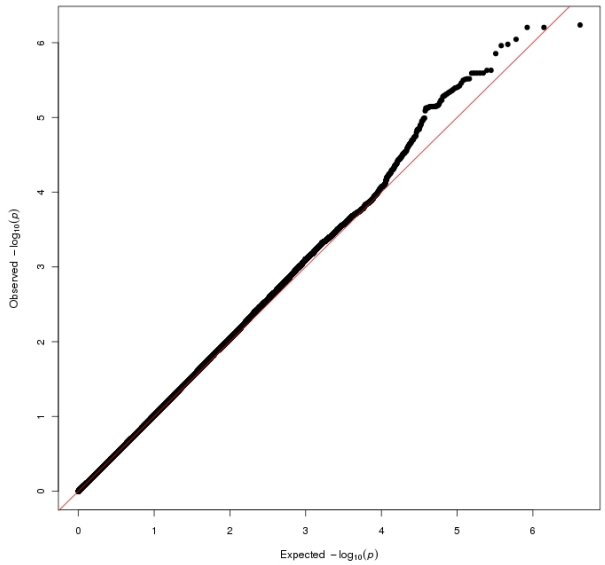
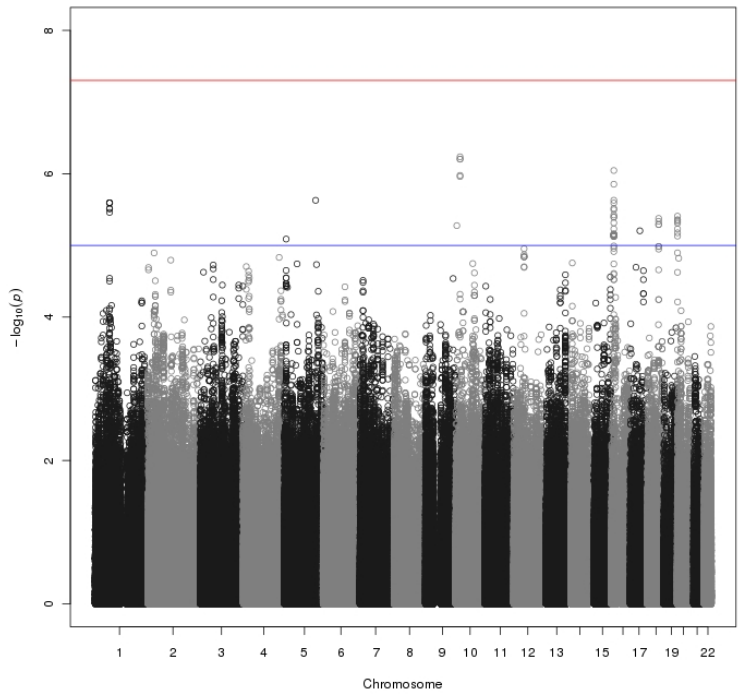


Figure 11. Manhattan plot of the NTR GWAS on BMI at the latest age between 24-120 months with SNPs cleaned on MAF > sqrt(5/1810) and Plink INFO > 0.4.



6. Genome-wide association study on pubertal staging

Within EGG, a GWA meta-analysis on pubertal staging using the Tanner stages was initialized. Within the NTR, data on Tanner stages have been collected in several projects. In some projects, Tanner staging was based on self-reports using pictures or drawings. In others, Tanner staging was assessed by a doctor or a trained researcher. Analyses were performed for each sex group separately and restricted to the age bins 10.5-12.5 years for girls and 12.6-15 years for boys. In the NTR, the number of boys with data in the specified age bin was 59, and it was decided that the sample size was too low to justify a GWA analysis. The number of Caucasian girls with GWA data and data on Tanner staging, was 265. The mean age at assessment was 12.06 (SD 0.32). Tanner stage was analyzed as a quantitative trait in Plink by linear regression with exact age at measurement and SNP dosage scores in an additive model as predictors. As the PCs were not significantly associated with the phenotype, none of them were included in the analysis. The --family option was used to correct for the relatedness of the data. One individual of each MZ twin pair was included, so that only full sibling relationships are present in the final dataset. The lambda of all results was 1.019 after cleaning on MAF > sqrt(5/265) and Plink INFO > 0.4. A Manhattan and QQ plot of the data cleaned on MAF and Plink INFO are shown in Figure 12 and 13.

A follow-up analysis was performed on the same dataset excluding girls with a BMI in the top 20% (n=188), as it was hypothesized that BMI interferes with a correct estimation of breast development in adolescent girls. A lambda of 1.042 was observed after cleaning on MAF > sqrt(5/188) and Plink INFO > 0.4. A Manhattan and QQ plot of the cleaned results are shown in Figure 14 and 15.

The results of both analyses have been uploaded to the server and the meta-analysis is currently in progress.

Figure 12. QQ-plot of the NTR GWAS on Tanner staging, SNPs cleaned on MAF > sqrt(5/265) and Plink INFO > 0.4.

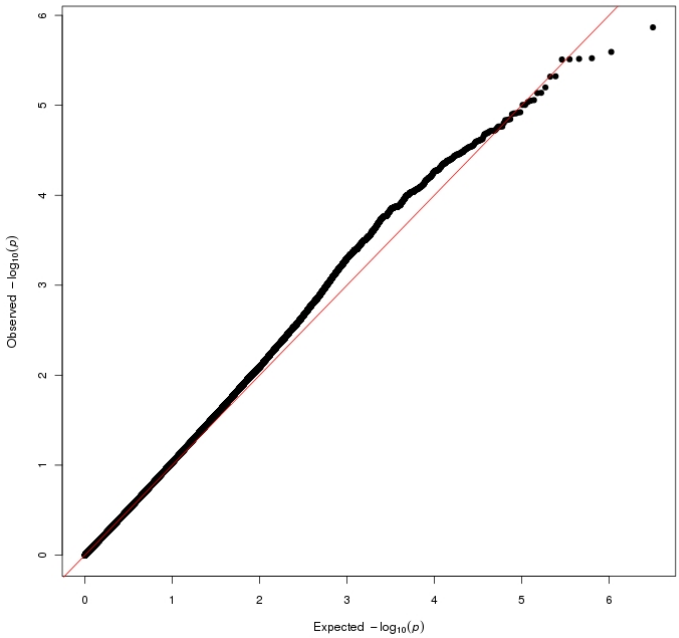


Figure 13. Manhattan plot of the NTR GWAS on Tanner staging, SNPs cleaned on MAF > sqrt(5/265) and Plink INFO > 0.4.

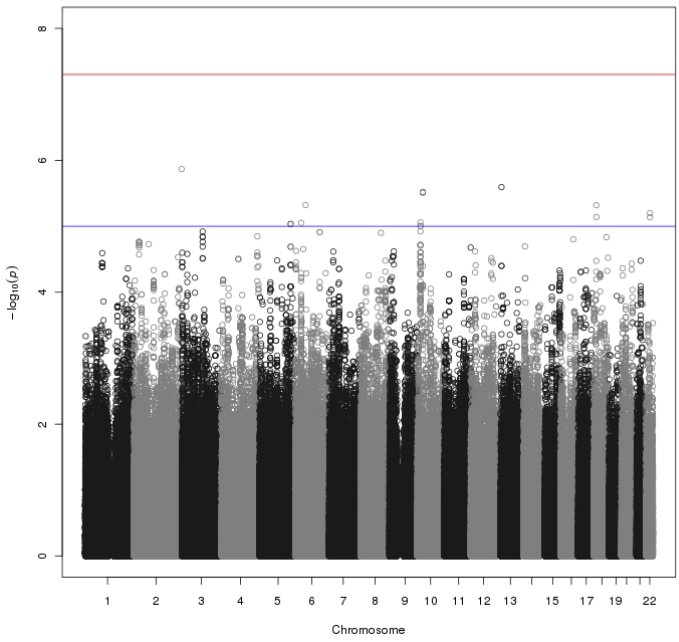


Figure 14. QQ-plot of the NTR GWAS on Tanner staging excluding girls with a BMI in the top 20%, SNPs cleaned on MAF > sqrt(5/188) and Plink INFO > 0.4.

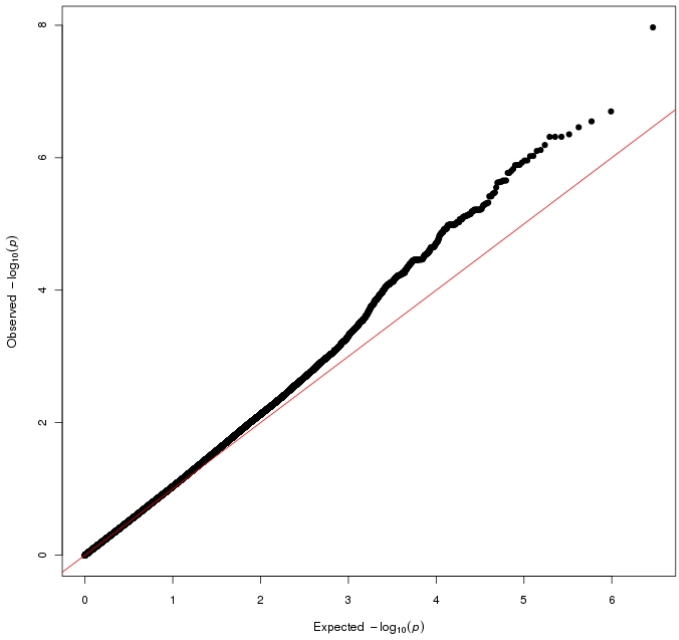
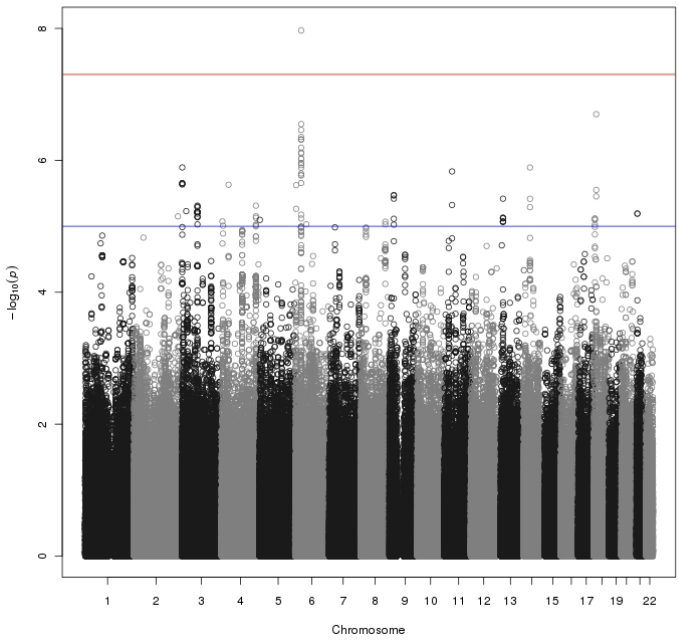


Figure 15. Manhattan plot of the NTR GWAS on Tanner staging girls with a BMI in the top 20%, SNPs cleaned on MAF > sqrt(5/188) and Plink INFO > 0.4.



7. Genome-wide association meta-analysis on height and growth in puberty

A GWA meta-analysis on height and growth in puberty was also conducted within EGG. Several analyses were conducted, focusing on prepubertal height (female height at 10, male height at 12) pubertal growth (age 8 to adulthood) and late pubertal growth (age 14 to adulthood). Analyses were performed for the two sex-groups separately and combined. The NTR participated late in the replication stage. Consequently, only the result of one SNP was used in the meta-analysis, namely rs281379 for the replication of a signal of male height at age 12. As described previously, data on height at age 12 were available from maternal reports (survey at age 12) and, for a smaller subgroup, laboratory measures. The measurement closest to the requested age within one year of the requested age was selected. The software package Growth analyser 3 (2004) was used to calculate sex specific age standardized scores using the 1997 Dutch reference growth charts for the general population. After exclusion of individuals with non-Dutch / non-European ancestry, data were available for 602 males. Plink was used to perform a linear regression in which age-adjusted height standard deviation scores were regressed on SNP and the PC that correlates with a north-south gradient in the Netherlands, as this PC was significantly correlated with the phenotype. Within the NTR, the effect observed in the discovery meta-analysis was not replicated for rs281379 (Table 2). In the final meta-analysis, 10 genome-wide results were found for the different phenotypes analyzed (6 for the combined analysis of male and female height at age 12/10).[5] Eight of those signals have previously been associated with adult stature, three with age at menarche and one with BMI. Only one SNP had not been implicated in any of these phenotypes.

Table 2. Results of the linear regression of height SDS scores at age 12 in males for the NTR and the discovery meta-analysis.

Cohort	**N**	**SNP**	**Chr**	**Effect allele**	**Non effect allele**	**Beta**	**SE**	**P-value**
NTR	602	rs281379	19	G	A	-0.04	0.063	0.5278
Discovery	4282	rs281379	19	G	A	0.08	0.017	2.85 E^{-6}

8. Genome-wide association study on motor development

Independent locomotion is considered an important developmental milestone; general motor development is associated with intelligence and delayed achievement is considered an early predictor of neuropsychiatric conditions.[6] Following a GWAS study on tooth development that found 5 genome-wide significant hits in a relatively small sample (n~6,000),[7] a GWA meta-analysis on motor development within EAGLE was proposed, specifically focusing on the age at which children can walk without support. Cohorts were asked to analyze the data using a survival model with a log-logistic distribution as this was previously shown to well describe the data of age at independent locomotion. In addition, such models deal with any interval or right censoring of the data that might occur in some cohorts. Children born prematurely (< 37 weeks) and children with cerebral palsy or intellectual disabilities (IQ < 70) were excluded from the analyses. One child of each NTR family was selected at random. After exclusion of individuals with non-Dutch / non-European ancestry, this led to a total sample size of 696 children. Log time to walking was regressed on sex, gestational age and genome-wide SNP data using the survival library in R (script provided by Alexandra Lewin). No PCs were included in the analyses as none were associated with the phenotype. In addition, the analyses were repeated for males and females separately. As some cohorts had no information on the exact age of locomotion but did have data on whether children were walking without support at either age 15 or 18 months, an additional set of analyses was run in which the data were dichotomized at those ages. The results for the NTR analyses are summarized in Figure 16-21. The lambda for each of the cleaned sets was 1.077 (survival regression), 1.018 (dichotomized at 15 months) and 1.015 (dichotomized at 18 months).

A preliminary meta-analysis has been performed for this project, but no genome-wide significant findings were found. Given the relatively small sample size of the total discovery set of the survival regression analysis (N ~6,600), the discovery phase has been reopened.

Figure 16. QQ-plot of the NTR genome-wide survival regression of age at walking without support filtered on MAF > sqrt(5/696) and Plink INFO > 0.4.

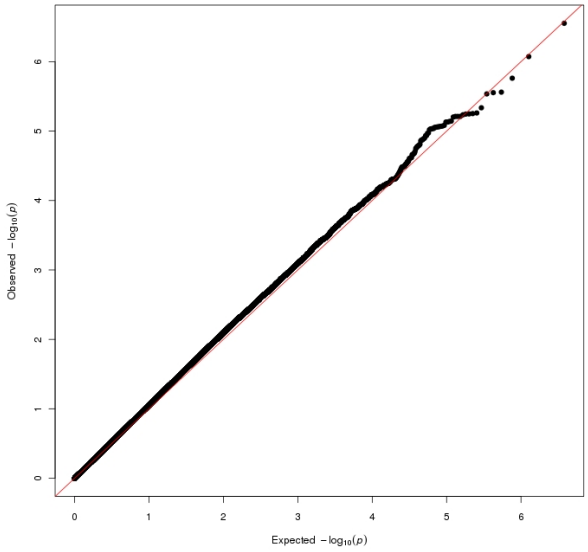


Figure 17. Manhattan plot of the NTR genome-wide survival regression of age at walking without support filtered on MAF > sqrt(5/696) and Plink INFO > 0.4.

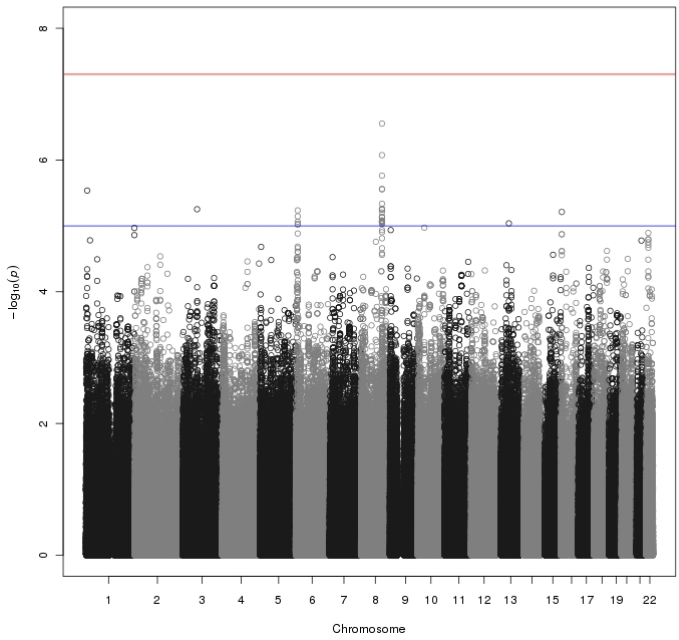


Figure 18. QQ-plot of the NTR genome-wide logistic regression of walking without support at age 15 months filtered on MAF > sqrt(5/696) and Plink INFO > 0.4.

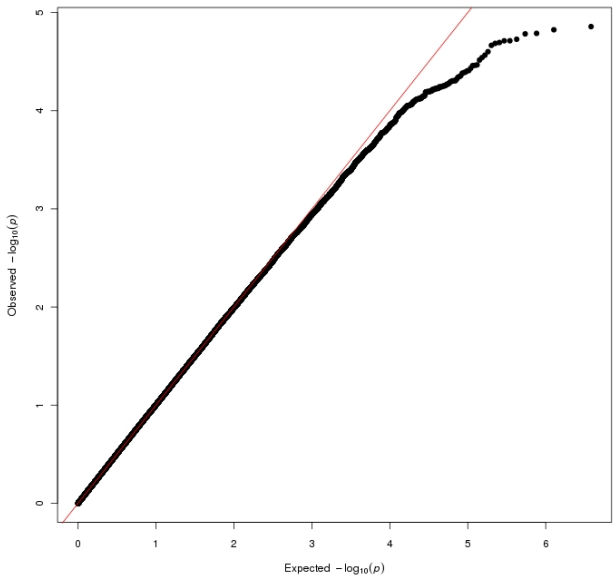


Figure 19. Manhattan plot of the NTR genome-wide logistic regression of walking without support at age 15 months filtered on MAF > sqrt(5/696) and Plink INFO > 0.4.

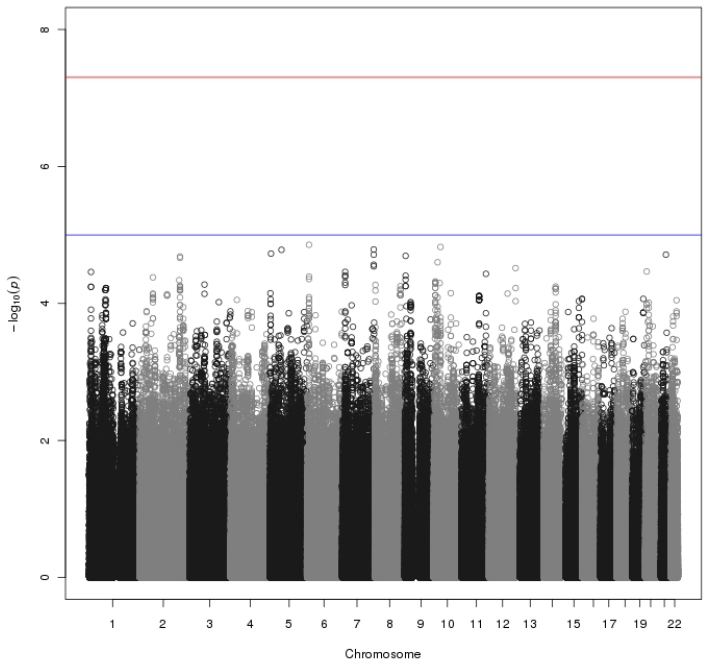


Figure 20. QQ-plot of the NTR genome-wide logistic regression of walking without support at age 18 months filtered on MAF > sqrt(5/696) and Plink INFO > 0.4.

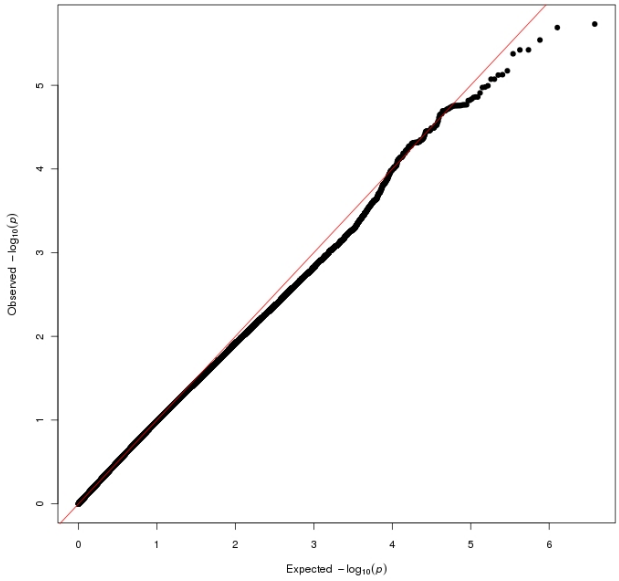
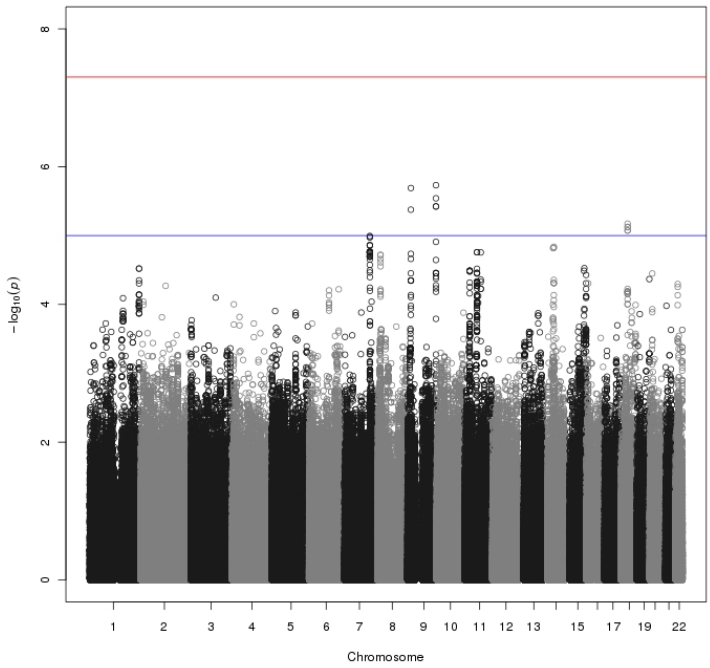


Figure 21. Manhattan plot of the NTR genome-wide logistic regression of walking without support at age 18 months filtered on MAF > sqrt(5/696) and Plink INFO > 0.4.



9. Genome-wide interaction study on birth weight and maternal smoking

Within EGG, a genome-wide interaction study was conducted in order to find genetic variants that moderate the effect of maternal smoking during pregnancy on fetal birth weight. Twins were excluded from the discovery phase of the study, but invited to take part in the replication phase. Fourteen SNPs that showed a p-value < $1E^{-4}$ in the discovery meta-analysis were selected for replication, as well as up to 5 proxy SNPs for each of the 14 lead SNPs. In the NTR, data on maternal smoking were based on a survey that is sent out after registration of the twins in the first year after birth. The way in which maternal smoking during pregnancy was assessed in this first survey has changed over time (see Table 3). The consortium requested to recode the data in a binary fashion with mothers that never smoked coded as 0 and mothers who smoked more than one cigarette per day at any time during pregnancy as 1 (see Table 3 for the recoding of each version of the NTR data). Only Caucasian, term-born (>= 37 weeks of gestation) children were included in the analyses, and one child of each family was selected at random, resulting in a sample size of 734 children. Birth weight data were quantile normalized using an R script provided by the consortium.

Analyses were performed in quicktest using the MRG4 dataset. One out of the 14 index SNPs and 6 out of the 65 proxy SNPs did not pass the SNP quality control filters applied after imputation and were excluded from the analyses. A linear regression was run in which quantile normalized birth weight was regressed on gestational age, sex, the 3 'Dutch' PCs, one dummy covariate indicating the chip on which individuals where genotyped, SNP dosage scores, maternal smoking during pregnancy and the interaction term of SNP dosage score by maternal smoking during pregnancy. The results are shown in Table 4 (13 index SNPs from the meta-analysis) and Figure 22 (13 index SNPs plus 59 proxy SNPs). In addition, a linear regression was run in which quantile normalized birth weight was regressed on gestational age, sex, the 3 'Dutch' PCs, the previously mentioned chip covariate and SNP dosage scores stratified by maternal smoking during pregnancy.

A combined meta-analysis of the discovery and replication phase is under way.

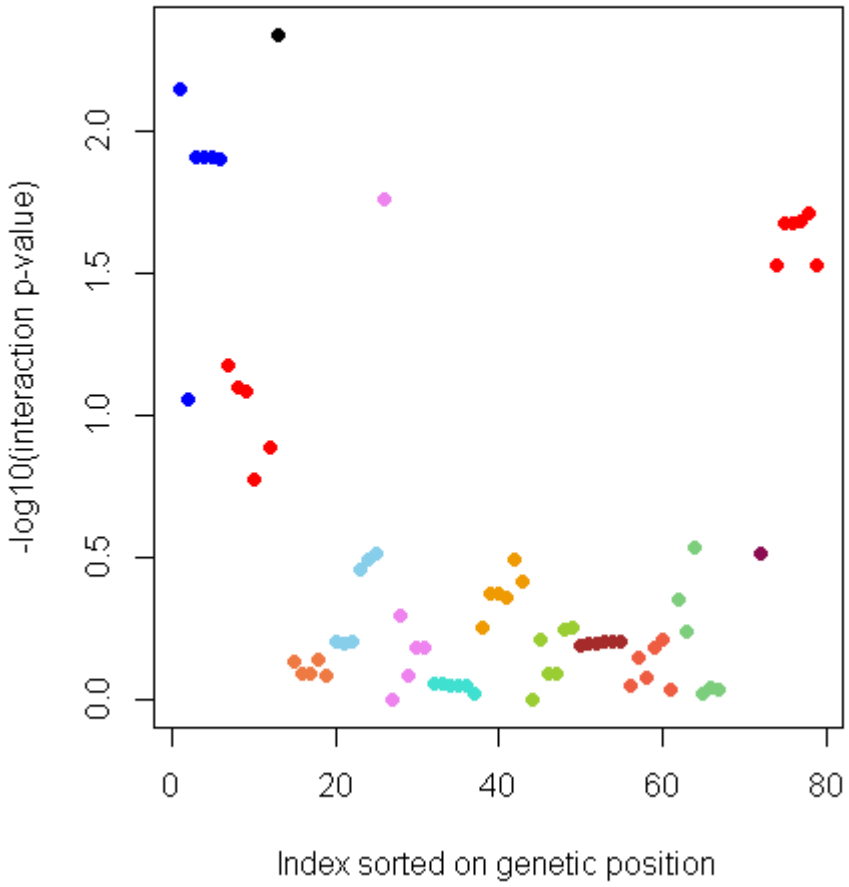
Table 3. Assessment of maternal smoking during pregnancy in the different NTR questionnaires over time and their recoding for the current analysis.

Version	Answer categories	Recoded as
1.	1. No, not at all.	0
	2. Yes, now and then	NA
	3. Yes, during the whole pregnancy	1
2.	1. no.	0
	2. Yes, pipe, cigars	NA
	3. Yes, less than 10 cigarettes per day	1
	4. More than 10 cigarettes per day	1
3.	1. No.	0
	2. Yes, pipe, cigars	NA
	3. Yes, 0-5 cigarettes per day	1
	4. Yes, 5-10 cigarettes per day	1
	5. Yes, more than 10 cigarettes per day.	1

Table 4. Results of the linear regression of quantile normalized birth weight on the interaction of the 14 top SNPs from the discovery meta-analyses with maternal smoking during pregnancy.

SNP	**Chr**	**Position (B36)**	**Gene**	**Description**	**Interaction p-value NTR**
rs4551142	6	77 922 500	NA	Inter-genic region	0.634
rs10825665	10	51 744 000	SGMS1	Sphingomyelin synthase 1	0.554
rs16977814	15	55 966 900	NA	Inter-genic region	0.445
rs11729139	4	63 021 300	NA	Near pseudo gene	0.734
rs3737907	1	7 646 120	CAMTA1	Calmodulin binding transcription activator 1	0.007
rs4779939	15	29 985 100	NA	Inter-genic region near OTUD7A and CHRNA7	0.881
rs2151877	10	109 820 000	NA	Inter-genic region	0.800
rs691605	8	19 072 900	NA	Inter-genic region	0.943
rs4140870	22	35 955 700	RAC2	Ras-related C3 botulinum toxin substrate 2	0.029
rs765161	2	215 483 000	NA	Inter-genic region	0.005
rs2906644	7	99 794 250	PILRB	Paired immunoglobin-like type 2 receptor beta	0.997
rs6435220	2	149 570 000	KIF5C	Kinesin heavy chain isoform 5C	0.080
rs6069125	20	53 159 625	NA	Pseudo gene	NA
rs8016733	14	26 628 625	NA	RNA gene	0.640

Figure 22. Plot of the negative log of the p-values of the linear regression of quantile normalized birth weight on the interaction of the 13 top SNPs from the discovery meta-analyses and 59 proxy SNPs with maternal smoking during pregnancy. Note that the SNPs are sorted on chromosome and position and that each cluster of a lead SNPs plus its proxy SNPs is plotted in the same color.



Conclusion

The projects reported here are in various stages; some are still waiting for cohorts to provide results for the discovery phase, other have finished their follow-up analyses and have resulted in published papers. Genome wide significant findings have been published for the meta-analyses on atopic dermatitis and pubertal height growth, several other projects have also identified variants at genome-wide significant levels in their preliminary analyses. These studies thereby show the power of collaboration in genetics, and the success of large-scale GWA studies. Moreover, they indicate that a specific meta-analysis of pediatric cohorts is worthwhile, as both new and known loci have been found for phenotypes such as height and BMI, phenotypes that were previously analyzed in adults only.

II. Project description, Aims, and Goals

The "Grand Opportunity" (GO) (American Recovery and Reinvestment Act (ARRA)) enabled genotyping in participants from the NTR. The project was financed by the NIMH and was entitled 'Genomics of Developmental Trajectories in Twins' (Grant: 1RC2MH089995-01; PI: James Hudziak).

In a later stage the project was named DETECT (Developmental Trajectories in Twins). The project involved collaboration of the NTR (VU University, Amsterdam) and, in the USA, the University of Vermont, the Avera Institute for Human Genetics and the University of Houston.

Goals

The goals of the project were formulated as follows:

1. Generate genome-wide Single Nucleotide Polymorphism (SNP) data and carry out GWA analysis to uncover associations between SNPs / haplotypes and behavioral and emotional problems in children. Associations will be tested using pedigree-based approaches and will identify SNPs that influence average trait expression across childhood.
2. Use SNP data from MZ twin pairs to carry out GWA tests for sensitivity genes that influence trait expression as a function of environment. We expect only partial overlap between genes influencing the average expression of a trait and genes influencing the sensitivity to environmental exposures.
3. Complete a GWA study on copy number variants (CNVs) in these same families. CNVs will be assessed in both members of MZ pairs, and in all other family members.
4. Compare the location and inheritance patterns of CNV across twin pairs who are concordant for having a behavioral phenotype, concordant for not having the phenotype, and discordant for the phenotype. In the entire sample, count total number of CNVs as a function of phenotype.
5. Identify MZ twin pairs who are discordant for CNVs; search the phenotype database for parallel discordance in behavioral and emotional problems. CNV's will play a greater role in the development of psychopathology in affected members of MZ pairs discordant for a trait (e.g. an MZ pair where one member has ADHD and the other does not) than in concordant affected pairs.

6. In a subset (N = 606 subjects) assess CNV in blood and buccal samples collected at the same time in order to determine across-tissue CNV structure.
7. In a second subset (N = 330; (165 complete twin pairs) carry out a longitudinal assessment of stability and changes in CNV in buccal samples assessed at two points in time.

III. Sample shipment and selection for DETECT

The hereafter reported quality control and imputation was sometimes performed on samples from the DETECT project only, and sometimes on samples from DETECT combined with samples from other NTR projects; where only samples from DETECT were used, this will be mentioned in the text.

Starting end 2009, samples collected by the NTR were shipped to the Avera Institute for Human Genetics (South Dakota, USA). In total 12,463 samples were shipped belonging to 7,419 individuals. For some individuals, multiple samples were available, that is, samples that were collected at different time points or from different tissues. Next to these multiple samples, the total of 12,463 samples includes 3,593 duplicate samples, meaning that one buccal DNA sample was taken at a particular time point and divided in two portions; the two portions were then counted as separate samples. Removal of those duplicate samples resulted in 8,870 samples belonging to 7,419 individuals. Most samples consisted of isolated DNA. In some cases, the samples consisted of buccal cells that still needed DNA isolation.

Most samples came from twins, either from children or from adolescent / adult twins (a large group of twins had provided buccal DNA for zygosity typing and for research purposes). Samples also came from parents of (young) twins, and others came from non-twins (e.g. siblings of twins). The year of sample collection ranged from 1994 to 2009.

Not all 8,870 samples were of equal relevance for the DETECT project and thus samples were prioritised for fingerprinting, which would provide information on zygosity and familial relatedness, and for genome-wide genotyping on the Affymetrix 6.0 platform. A further selection was needed as financial resources were available for genome-wide genotyping of ~4,700 samples, this selection was based on the availability of highly informative phenotypes for DETECT and included participants that were registered with the Young NTR (YNTR) or with the Adult NTR (ANTR).

Samples were prioritized in the following order:

1. All persons with blood and/or buccal samples of twins, sibs and parents who participated in the study of Dr. Hoekstra and Dr. Estourgie-van Burk and were followed from age 5 in a project in

which longitudinal IQ, EEG and other data were collected.[8, 9] In the lab, these samples were known as the "ROSA" group.

2. All persons with blood and/or buccal samples of the CHIMEAR group. This study collected blood samples at a young age.[10]
3. All persons with longitudinal DNA samples from buccal cells who met the following criteria: - 2 year or 7/8 year period between buccal collection.
   - both twins have either 2 or 7/8 year follow-up.
   - in case a sibling was present, his/her buccal sample was included
   (2 samples if 2 or 7/8 year follow-up present).
   - in case parents were present, parental buccal samples were included
   (only one buccal sample per parent).
4. Twins selected to be informative for ADHD whose mothers had participated in the DISC interview for their children.[11]
5. Twins and siblings with psychometric IQ data (multiple projects; e.g. projects by Dr. van Baal, Dr. Bartels, Dr. van Soelen, Dr. Polderman and Dr. van Beijsterveldt).
6. Twins and siblings with MRI data (projects by Dr. van Leeuwen, Dr. Peper and Dr. van Soelen).
7. Participants with many relevant parental and teacher ratings of behavioural phenotypes (e.g. CBCL and TRF).
8. Participants with many relevant self-ratings of behavioural phenotypes (YSR, YASR, ASR).
9. Participants with few relevant YNTR phenotype data.
10. Participants with few relevant ANTR phenotype data.
11. Participants with no relevant phenotype data.

These priority codes are consecutive – all twins in the first group have IQ data but are not counted again for priority 5. For some individuals, the priority scores changed, because they had already been genotyped in previous or ongoing studies. This group was manually examined, as family structure was also important for priority 1 and 2. Next, for the persons with priority code of 4 or higher, one sample (if duplicates or multiples were available) was selected for the analysis. This selection was based on the most recent sample of isolated DNA.

Out of the total of 8,870 samples, 2,573 were already genotyped and 468 did not fulfil the criteria to be included as a longitudinal sample. This resulted in a total of 5,829 samples with a priority code from 1-10 representing 5,034 individuals from 1,904 families. All these were selected to be fingerprinted on a 64 SNP fingerprint chip. In a later stage, it was decided to exclude samples from the CHIMAER project as material was sparse and Affymetrix 6.0 genotyping was considered to be of greater importance. Of the 5,034 individuals, 680 had 2 samples, 56 had 3 samples and 1 person had 4 samples. For genotyping on the Affymetrix 6.0 platform, samples were selected if they had a priority code ranging between 1 and 9 (so individuals with few relevant YNTR data were included but individuals with few relevant ANTR data were not), leading to a total of 4,813 samples selected for genotyping on Affymetrix 6.0.

IV. Fingerprinting

As a subproject within DETECT, a large proportion of samples that were planned to be genotyped on Affymetrix 6.0 were also genotyped for 64 SNPs (on a fingerprint chip) and 5 VNTRs (Variable Number Tandem Repeats). The 5 VNTRs and 38 of the SNPs were selected from the literature as polymorphisms in candidate genes for ADHD and other behavioral phenotypes. The remaining 26 SNPs were selected from the Affymetrix 6.0 platform (see Table 6).

After the DETECT project was finished, different sets of SNPs, partly overlapping with the DETECT set, were used in ongoing projects for determination of zygosity and familial relatedness.

Four different sets of SNPs/VNTRs have been used over time;

- Set 1: a set of 64 SNPs (38 in candidate genes) and 5 VNTRs used in the DETECT project as described above, genotyping took place at the Avera Institute for Human Genetics
- Set 2: a set of 63 (37 in candidate genes) SNPs used after DETECT, genotyping took place at the VU medical centre
- Set 3: a set of 62 (36 in candidate genes) SNPs and 7 VNTRs, genotyping took place at Avera Institute for Human Genetics as a follow up on the genotyping project at the VU medical centre
- Set 4: a set of 30 SNPs (all in candidate genes) currently in use at the Avera Institute for Human Genetics

An overview of the number of samples and information on the individuals and family relations present in each file are provided in Table 5.

Table 5. Overview of the number of samples and individuals genotyped in each fingerprint set (before cleaning), as available December 2012.

N	**Set 1**	**Set 2**	**Set 3**	**Set 4**
Samples	5448	1152	365	456
Individuals	4702	1148	365	455
Families	1657	395	195	223
Twins	2899	699	310	353

Triplets	25	0	0	0
Siblings	267	56	43	64
Parents	1511	393	12	38

An overview of the different sets of SNPs and VNTRs that were used in each project is presented in Table 6.

Table 6. Overview of the different SNPs and VNTRs used in subsequent sets for determination of zygosity and familial relatedness (1=yes, 0=no).

Candidate Genes	**dbSNP ID#**	**Genotyped in Set 1**	**Genotyped in Set 2**	**Genotyped in Set 3**	**Genotyped in Set 4**
BDNF	rs6265 *	1	0	0	1
COMT	rs4680	1	1	1	1
DRD2	rs1800497	1	1	1	1
DBH	rs1611115	1	1	1	0
DBH	rs2519152	1	1	1	1
SNAP-25	rs3746544	1	1	1	0
SNAP-25	rs1051312 *	1	0	0	1
NET	rs998424	1	1	1	0
NET	rs3785157	1	1	1	1
CLOCK	rs1801260 *	1	0	0	1
ApoE	rs7412	1	1	1	0
ApoE	rs429358	1	1	1	1
NPY	rs16139	1	1	1	1
ADRA2A	rs1800544	1	1	1	1
ADRA2A	rs1800545	1	1	1	0
ADRA2A	rs553668	1	1	1	0
AADAT	rs13145318	1	1	1	1
PCLO	rs2715148	1	1	1	1

PCLO	rs2522833 *	1	0	0	0
HTR1B	rs6296	1	1	1	1
HTR2A	rs6314	1	1	1	1
TPH2	rs1843809	1	1	1	0
TPH2	rs1386497 *	1	0	0	1
GRIN2A	rs8049651	1	1	1	1
DRD1	rs265981 *	1	0	0	1
FADS2	rs174575	1	1	1	1
FADS2	rs1535	1	1	1	1
ApoE	rs2075650 *	1	0	0	1
CLU/ApoJ	rs11136000	1	1	1	1
PICALM	rs3851179	1	1	1	1
CR1	rs6656401	1	1	1	1
DAT	rs2652511	1	1	1	0
DAT	rs40184	1	1	1	1
DRD4	rs1800955 *	1	0	0	1
DRD4	rs3758653	1	1	1	0
SRC1	rs11125744 *	1	0	0	1
FTO	rs9939609	0	1	1	1
FTO	rs1121980	0	0	0	1
FTO	rs6499640	0	0	0	1
FTO	rs8050136	0	0	0	1
ADRB2	rs1042714	0	1	1	0
CHRNA5	rs16969968	0	1	1	0
FADS1	rs174550	0	1	1	0
GRM8	rs2237781	0	1	1	0
GRIK2	rs6570989	0	1	1	0
CD53	rs6679497	0	1	1	0
GABRG3	rs8036270	0	1	0	0

Remaining SNPs	dbSNP ID#	Genotyped in Set 1	Genotyped in Set 2	Genotyped in Set 3	Genotyped in Set 4
SNP on Y chromosome	rs768983				
Affymetrix SNP	rs10010369	1	1	1	0
Affymetrix SNP	rs11123433	1	1	1	0
Affymetrix SNP	rs1333340	1	1	1	0
Affymetrix SNP	rs1709915	1	1	1	0
Affymetrix SNP	rs17272796	1	1	1	0
Affymetrix SNP	rs1810467	1	1	1	0
Affymetrix SNP	rs2128347	1	1	1	0
Affymetrix SNP	rs216459	1	1	1	0
Affymetrix SNP	rs2306641	1	1	1	0
Affymetrix SNP	rs242076	1	1	1	0
Affymetrix SNP	rs2499636	1	1	1	0
Affymetrix SNP	rs2577256	1	1	1	0
Affymetrix SNP	rs2588333	1	1	1	0
Affymetrix SNP	rs4775699	1	1	1	0
Affymetrix SNP	rs4793172	1	1	1	0
Affymetrix SNP	rs6469617	1	1	1	0
Affymetrix SNP	rs6734275	1	1	1	0
Affymetrix SNP	rs685449	1	1	1	0
Affymetrix SNP	rs6909592	1	1	1	0
Affymetrix SNP	rs7116710	1	1	1	0
Affymetrix SNP	rs7627201	1	1	1	0
Affymetrix SNP	rs7728335	1	1	1	0
Affymetrix SNP	rs7795959	1	1	1	0
Affymetrix SNP	rs7924984	1	1	1	0
Affymetrix SNP	rs9293511	1	1	1	0

Affymetrix SNP	rs9865763	1	1	1	0
VNTRs	**Polymorphism**	**Genotyped in Set 1**	**Genotyped in Set 2**	**Genotyped in Set 3**	**Genotyped in Set 4**
SERT	5-HTTLPR + rs25531	1	0	1	0
DRD4		1	0	1	0
DAT1		1	0	1	0
DRD5		1	0	1	0
MAOA		1	0	1	0
DRD4	Promotor, 120 bp-repeat	0	0	1	0
SERT	Intron 2, 17-bp repeat	0	0	1	0
Amelogenin	6 bp deletion on AMELX	1	1	1	0

Within the DETECT set (Set 1), nine SNPs did not perform well initially; these are marked with an asterisk in Table 6. All nine SNPs were successfully rerun on the Taqman if samples had sufficient material available; this excluded 162 buccal and 382 blood samples (a total of 4,904 samples were retyped). The five VNTRs were genotyped only in the set of 4,904 samples with sufficient material. The nine SNPs that were retyped replaced the original typings. For samples that were not retyped, the nine SNPs were put to missing.

QC was applied to each set separately. SNPs and VNTRs were tested for call rate, HWE, and correspondence of the allele frequency in the current set with the frequency in the reference set (HapMap release 22, CEU panel) In some sets there were enough parent-child relationships to test for Mendelian errors and in the DETECT set, a substantial number of individuals had multiple samples genotyped allowing for a test of concordance of SNPs across duplicate samples. Table 7 lists the SNPs that were excluded in each subset with the reason for exclusion.

Table 7. SNPs showing problematic QC.

	SNP	ME	Blood-buccal errors	HWE
Set 1	rs7412	Problematic	Problematic	Not in HWE
	rs2652511	Problematic	Problematic	In HWE
	rs429358	Problematic	Problematic	Not in HWE
Set 2	rs6469617	OK	Not applicable	Not in HWE
	rs7795959	OK	Not applicable	Not in HWE
	rs2519152	OK	Not applicable	Not in HWE
	rs7412	Problematic	Not applicable	Not in HWE
Set 3	rs429358	Not applicable	Not applicable	Not in HWE
Set 4	To be done	To be done	To be done	To be done

Across sets, samples were checked for Mendelian errors, sex errors, inconsistencies between blood-buccal and longitudinal samples and inconsistencies between questionnaire and genotype-based family relations and zygosity, where possible. Samples that were suspected sample switches were put on a blacklist and excluded from further analyses. The call rate of each sample was determined separately for SNPs and VNTRs, all SNP calls were put to missing if the call rate was < 90%, for the VNTRs, all VNTR genotypes were put to missing if more than one was missing. Table 8 gives the number and percentage of samples that were excluded from each subset. Note that the total number of excluded samples is sometimes lower than the number of samples with e.g. a low SNP call rate, due to the fact that some samples had a low call rate for the SNPs but not the VNTRs and vice versa.

Table 8. Number of samples showing problematic QC.

	Set 1	Set 2	Set 3	Set 4
N total	5,448	1,152	365	456
N on blacklist	115 (2.1%)	9 (0.8%)	0 (0%)	To be done
N with low call rate SNPs	201 (3.7%)	50 (4.3%)	11 (3.0%)	4 (0.9%)
N with low call rate VNTRs	38 (0.8%)	Not applicable	12 (3.3%)	Not applicable
N excluded	155 (2.8%)	59 (5.1%)	5 (1.4%)	4 (0.9%)

Note that some individuals were genotyped multiple times within or across subsets, leaving the total number of individuals that have a genotyped sample that passed QC of either fingerprint or VNTR data at 6,347. A final file was created in which each individual has only one genotype call per SNP/VNTR. Only calls from non-blacklisted samples that are consistent across samples from the same individual were included in the final dataset.

Zygosity was determined using the program ZygProb/ECLIPSE2 (available via: http://gump.qimr.edu.au/general/daleN/ZygProb/).[12, 13] This program estimates the MZ/DZ likelihood for each twin pair at different genotyping error rates. For the current dataset, a genotyping error rate of 0.01 was chosen. In case the genotype-based zygosity differed from the survey-based zygosity, the questionnaire and genotyped data were checked to see whether the observed difference was likely due to a sample swap or low quality sample or to an incorrect assignment of questionnaire based zygosity.

Twin families who had indicated that they wanted to be informed about their zygosity received a letter with the results of the zygosity determination. In the letter they were asked to contact the NTR if there were reasons to question the zygosity assessment.

V. Affymetrix Genotyping

Microarray processing. SNP and CNV data were generated from each DNA sample using high-density SNP 6.0 microarrays (Affymetrix; Santa Clara, CA, USA). This array represents 1.8 million markers evenly distributed across the whole genome, including more than 906,600 markers to detect SNPs and another 946,000 probes for the detection of CNVs.

Sample preparation and microarray processing were performed according to Affymetrix protocols. In brief, 500 ng of purified genomic DNA is prepared for restriction digest, according to strict protocol specifications. Two aliquots (250 ng) of each sample were used. One was digested by the restriction enzyme NspI, the other by StyI (New England Biolabs; Ipswich, MA, USA). The digested samples were then subjected to adaptor-ligation using T_4 DNA ligase (New England Biolabs). The adaptors recognize the cohesive 4bp overhangs; consequently, all fragments from the digest are substrates for ligation, regardless of size.

A polymerase chain reaction (PCR) was used to amplify the genomic DNA for each of the ligated samples, using a generic primer which recognizes the universal adaptor sequence. Assay conditions have been optimized to generate fragments between 200bp and 1100bp. The optimal fragment size was confirmed by performing a QC 2% agarose gel electrophoresis. PCR products were then purified using an AMPure® magnetic bead protocol (Beckman Coulter Genomics; Danvers, MA, USA). Purified PCR product was quantified using a Spectramax® Plus 384 Spectrophotometer (Molecular Devices; Sunnyvale, CA, USA) to ensure adequate PCR fragment yield. The resulting DNA was subjected to fragmentation using the Affymetrix supplied DNA fragmentation enzyme (DNaseI) to produce fragments of optimal size for hybridization to the markers on the microarray (< 180bp). Following fragmentation, the digestion product size was confirmed using an additional QC check with a 4% agarose gel electrophoresis.

The resulting fragments were end-labeled (biotinylated) using an Affymetrix supplied DNA labeling reagent consisting of Biotin and the enzyme terminal deoxynucleotidyl transferase (TdT). The labeled products were mixed with a hybridization solution and denatured on a thermal cycler. The hybridization reactions were loaded onto a Genome-Wide Human SNP Array 6.0 and hybridization occurred for 16-18 hrs at 50ºC with 60 Revolutions Per Minute (RPM) of rotation.

Following hybridization, the arrays were subjected to the following cycles of staining and buffer stringency washes: a Steptavidin Phycoerythrin (SAPE) stain cycle, an Antibody Stain cycle, a second Streptavidin stain step, a final buffer wash, and finally the microarray is filled with holding buffer. All of the wash and stain steps were carried out in automated fluidics stations (Affymetrix; Santa Clara, CA, USA). The arrays were then scanned in a GeneChip Scanner 3000 7G (Affymetrix; Santa Clara, CA, USA) with an autoloader accessory. The command console software creates a (.DAT) file for each sample/array that is subsequently scanned. This file is captured by Affymetrix Genotyping Console 4.0, converted to the standard intensity file (.CEL) where it can be stored and/or converted into different files for further analysis.

All raw intensity (.CEL) and attribute (.ARR) files were uploaded to the Secure File Transfer Protocol server initially with no cleaning from Avera Institute for Human Genetics.

Genotype calling. Genotype calling from the Affymetrix CEL files was performed by Xiao Xiangjun and Paul Scheet at the University of Houston using Birdsuite version 2 (Birdseed).

CNV calling. CNV calling from the Affymetrix CEL files has been performed by Abdel Abdellaoui and Jouke-Jan Hottenga at the VU University.

VI. Imputation:

Between February 2011 and May 2013, five different imputations had been performed that included the DETECT Affymetrix 6.0 data. The first two included only DETECT data and were imputed against HapMap release 22, the third and fourth included all available NTR and NESDA samples in December 2011 and December 2012 respectively, and were imputed against the 1000 Genomes reference set. The fifth imputation included the same set of individuals as the fourth but including only one individual from each MZ pair, and was imputed against the GoNL reference set. The set that was used in the second round of imputation consisted of the full set of DETECT samples; for this set of samples the data cleaning will be described more extensively.

1. Imputation DETECT February 2011

In the beginning of 2010, an imputation was performed in Houston by Paul Scheet and Xiangjun Xiao on the first 26 Affymetrix 6.0 plates in order to use these data for replication of GWAS signals for consortia that were closing their analyses very soon afterwards. In total, there were 2,281 genotyped samples (excluding control samples), of which 1,603 were included in the imputed dataset. This is due to the fact that all individuals with multiple samples were genotyped on the first plates, and only one sample per individuals was included for imputation (the sample with the best quality was picked). In addition, 102 samples were excluded as they were suspected sample switches in the fingerprint set.

Prior to imputation, SNPs were filtered based on the following criteria: missing data rate, minor allele frequency (MAF), Hardy-Weinberg equilibrium (HWE) and Mendelian errors (Table 9). SNPs that passed all 4 criteria were propagated to the imputation stage. Since for the imputation program individuals are essentially independent, an individual-level filter was not applied. That is, poor quality of one sample will not adversely affect the imputation accuracy for other samples.

Table 9. Criteria for SNP cleaning and the number of excluded SNPs before imputation.

Criterion	**Value**	**N SNPs failing**
Call rate	< 0.90	1,275
MAF	< 0.01	101,191

HWE (p-value)	< 0.0001	4,134
MEs	> 35	1,171

There were 872,242 SNPs before the QC process. Of these, 761,750 pass all 4 criteria.

Imputation was conducted with the software BEAGLE using HapMap release 22 haplotype data of the CEU population as a reference set. BEAGLE was chosen for its ability to handle trios, which can be arranged and constructed to take advantage of the informative data structure in the samples from the NTR.[14] For purposes of speed and testing, in this initial phase, BEAGLE was run treating all individuals as unrelated. This has been shown to be reasonably accurate.

Post imputation samples were checked for call rate, family structure, Mendelian errors and heterozygosity. Samples that failed QC were excluded from the analyses.

2. Imputation DETECT September 2011

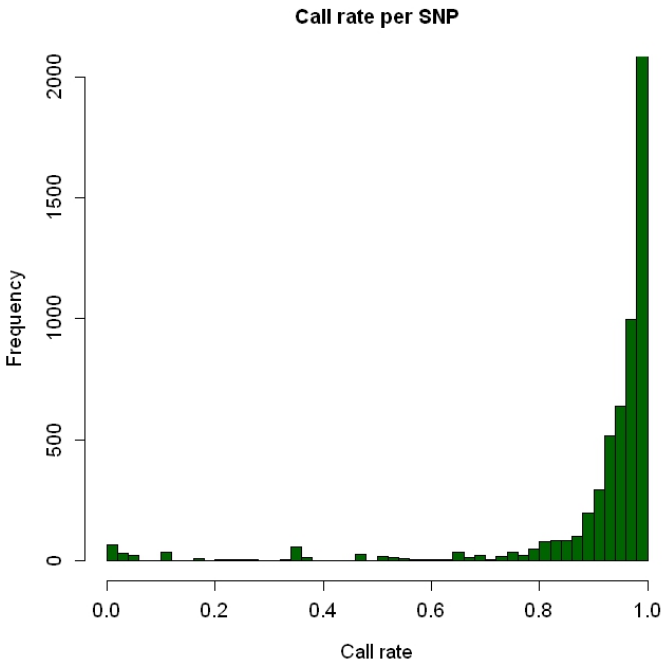
When all plates were genotyped, genotype calling and imputation on the total set of 55 plates and 4,801 samples were performed in Houston by Paul Scheet and Xiangjun Xiao. Genotype calling was again performed using Birdsuite version 2 (Birdseed).

Before imputation, the following sample selection was performed in Houston. Of the total of 4,801 samples, 169 were removed because they did not survive QC in the lab, the fingerprint dataset or in the first DETECT imputation. In addition, 110 samples were excluded because their Contrast Quality Control (CQC) was below 0.4 (cutoff suggested by Affymetrix). Duplicates were removed by selecting the sample with the highest quality, and one sample of each monozygotic twin pair was selected at random (for a twin pair thought to be monozygotic, one sample was removed if the IBD estimate in Plink was > 1.5, for a twin pair thought to be dizygotic, one sample was removed if the IBD estimate in Plink was > 1.8). This resulted in a total of 3,266 samples before imputation.

To ensure quality control, SNPs were checked for their call rate, minor allele frequency, Hardy-Weinberg equilibrium, Mendelian error rate and concordance rate in duplicate samples. Here, we will give a short, mainly graphical, summary of the results.

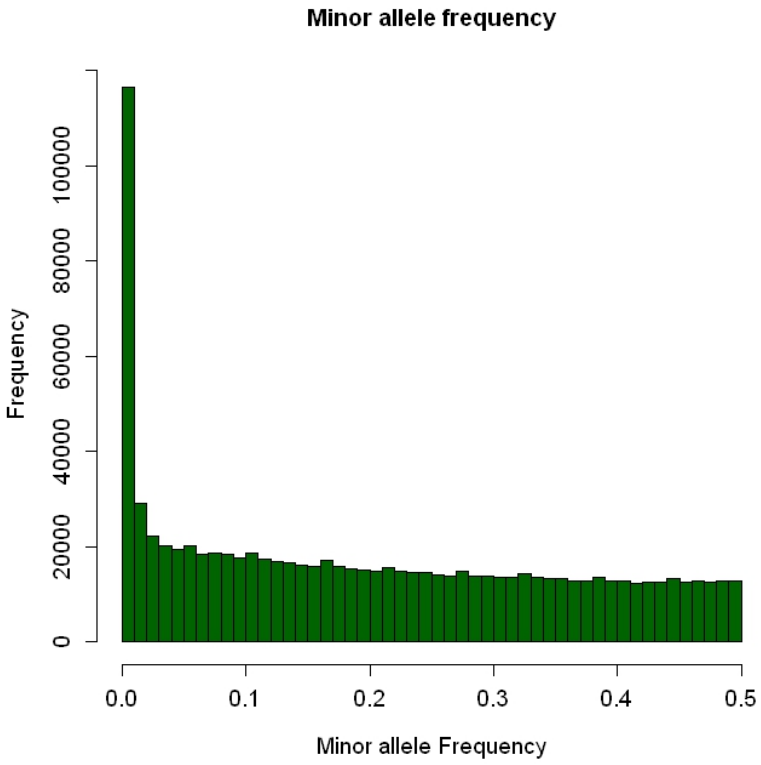
In Figure 23, a histogram of the SNP call rate is shown (note that the y-axis has been truncated at 2,000). The call rate is an important indicator of the quality of the genotype data. Only a few SNPs showed a call rate lower than 0.95 (2,379 out of 872,242 SNPs).

Figure 23. Call rate per SNP in the 3,266 samples included in the second DETECT imputation, genotyped SNPs (y-axis truncated at 2,000).



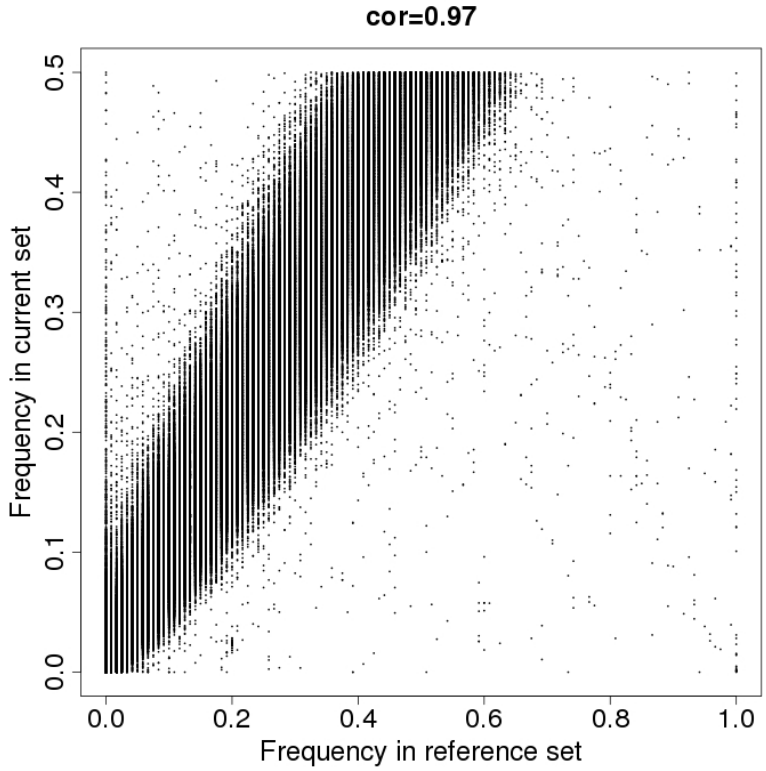
In Figure 24, the minor allele frequency of all SNPs is shown as calculated on the set of parents in DETECT (n=709). As expected, there is a relatively large number of low frequency SNPs, whereas the remaining SNPs shows a more or less uniform distribution. The quality of SNPs with a low minor allele frequency is generally worse, therefore SNPs with a MAF < 0.01 were excluded before imputation.

Figure 24. Minor allele frequency of the genotyped SNPs in709 parents in the DETECT set.



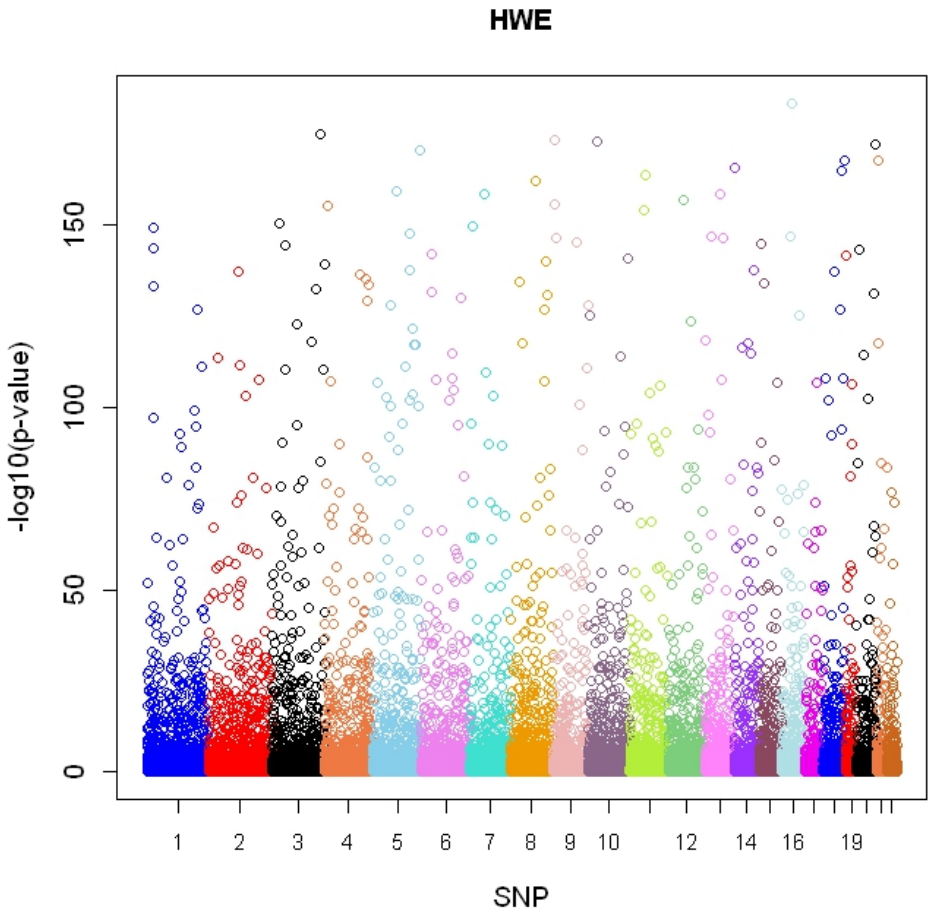
In Figure 25, the minor allele frequency in the DETECT set is plotted against the frequency of the same SNP and allele in a reference set (HapMap release 24, CEU reference panel), which showed a correlation of 0.97. The vertical lines in the left and right side of the plot consists of SNPs that were non-polymorphic in the reference set but polymorphic in the DETECT set.

Figure 25. Frequency of genotyped SNPs against the frequency of the same SNP and allele in HapMap release 24 CEU reference panel.



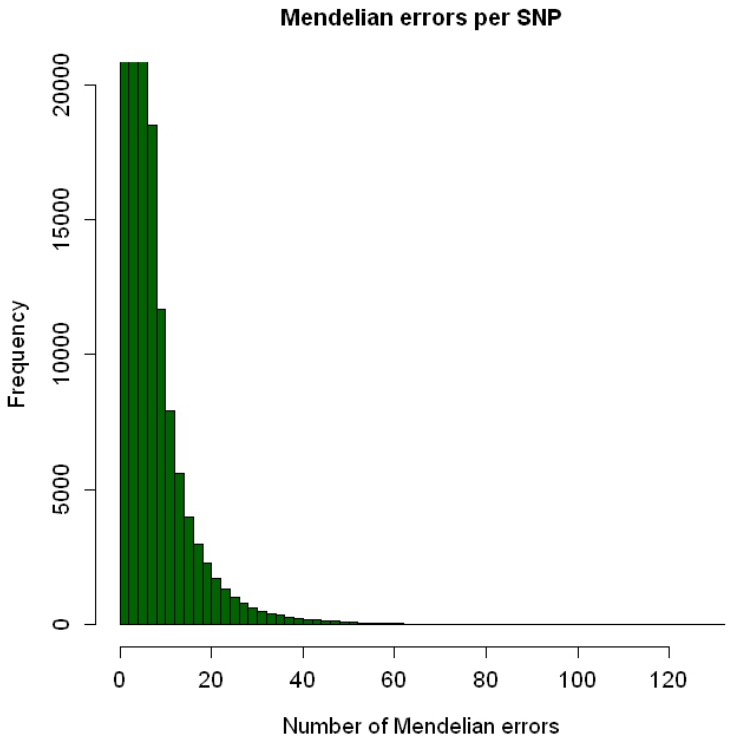
A common problem that occurs with genotype calling is that only two of the three possible genotype calls are being called. These kind of errors and other issues are readily identified by a test for Hardy-Weinberg equilibrium. In Figure 26, the minus log10 of the p-value of the test for Hardy-Weinberg equilibrium of all SNPs is plotted against their position in the genome. Strong deviations occurred across the genome; 4,245 SNPs showed a HWE p-value lower than $1E^{-5}$ and were excluded from the dataset before imputation.

Figure 26. Minus log10 of the p-value of the Hardy-Weinberg equilibrium test for all genotyped SNPs.



As the dataset consisted of family members, it is possible to test whether SNPs show an excess of Mendelian errors, another measure of SNP quality. A histogram of the number of Mendelian errors per SNP is show in Figure 27; SNPs with more than 26 Mendelian errors were excluded before imputation.

Figure 27. Histogram of the number of Mendelian errors per SNPs for the genotyped set (y- axis truncated at 20,000).



A check for the concordance rate of SNPs in duplicate samples was performed in Houston, leading to the exclusion of 24,573 SNPs with a concordance rate < 0.95. Table 10 gives an overview of the criteria that were used to clean the SNPs before imputation and the number of SNPs that were excluded based on each criterion.

Table 10. Criteria for SNP cleaning and the number of excluded SNPs before imputation.

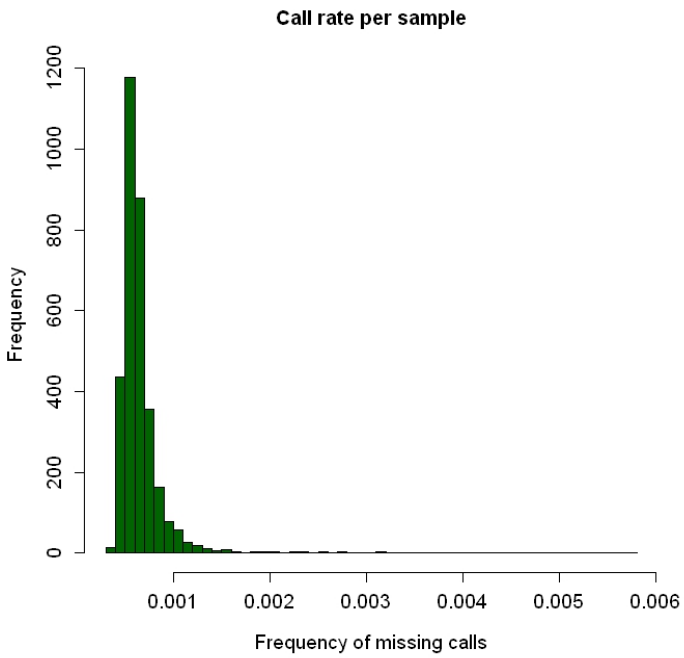
Criterion	**Value**	**N SNPs failing**
Call rate	< 0.95	2,379
MAF	< 0.01	Unclear
HWE (p-value)	< 0.00001	4,245
MIs	> 26	8,594
Concordance rate in duplicates	< 0.95	24,573

In total 726,842 of 872,242 genotyped SNPs passed all criteria and were used for imputation.

Imputation was performed with the software BEAGLE using HapMap release 22 haplotype data of the CEU population as a reference set. All individuals were treated as unrelated.

Post imputation sample cleaning was performed; all samples were checked for call rate, family structure, Mendelian errors, sex errors and heterozygosity where possible. As stated earlier, samples were already filtered in Houston based on CQC and previous quality checks. Figure 28 shows the call rate of the 3,266 samples that were imputed. The sample selection resulted in a very high call rate for the remaining samples, and, therefore, no additional samples were excluded based on this criterion.

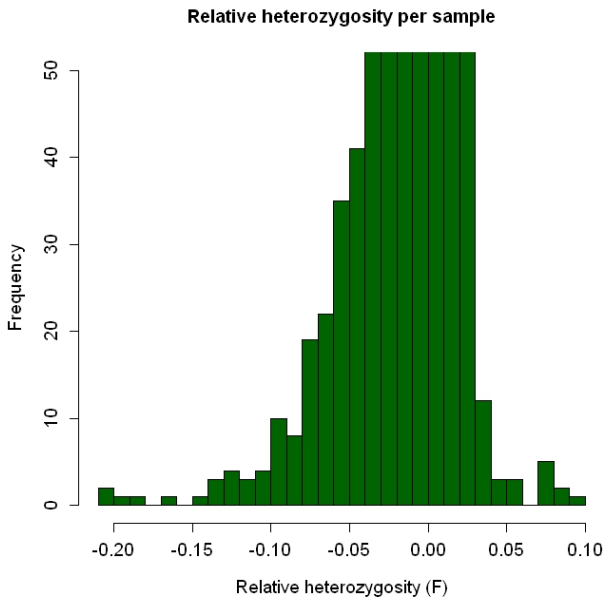
Figure 28. Histogram of the sample call rate for the 3,266 samples included in the second DETECT imputation.



Further checks were based on a dataset of genotypes that underwent basic cleaning (filtered on SNP call rate >= 0.95, SNP Mendelian errors <= 26, HWE p-value > 0.00001, MAF in DETECT and reference set > 0.01, difference of frequency with reference set < 0.1% and > 99.9% quantile).

As an excess of heterozygous calls can indicate sample contamination from e.g. surrounding samples on the plate, the relative heterozygosity of a sample is an important measure of its quality. A histogram of a measure of relative heterozygosity (F) estimated in Plink is shown in Figure 29. Samples with .10 < F < -.10 were excluded from further GWAS analyses on the imputed set (n=23). Another way to detect sample contamination and to check for sample switches is to investigate whether the known sex of the individual is in line with the genotypes on the X-chromosomes. This check was performed in Houston and led to the exclusion of another 33 samples.

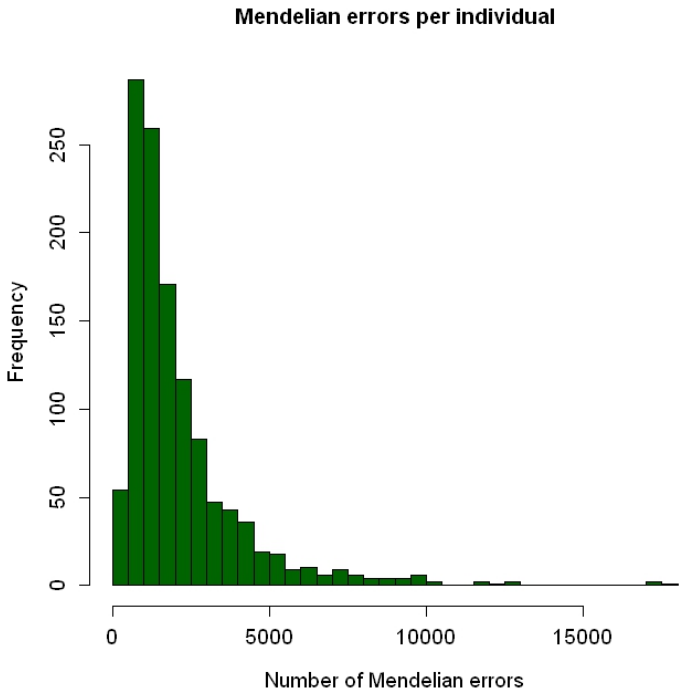
Figure 29. Estimated inbreeding coefficient from Plink, for the 3,266 samples included in the second DETECT imputation (y-axis truncated at 50).



Samples that showed genetic relations that were not in line with the expected relationships had already for a large part been identified during the quality control steps performed on the fingerprint dataset and the GWAS set imputed in February 2011, and were filtered out before imputation. A plot of the Mendelian errors per individual for the remaining 3,266 samples is shown in Figure 30. As the total number of possible Mendelian errors depends on the number of SNPs genotyped and the number of family relations observed for an individual, these numbers cannot be readily interpreted, but none of

the samples showed an Mendelian error rate > 2%, and therefore no samples were excluded based on this criterion.

Figure 30. Number of Mendelian errors per individual, for the 3,266 samples included in the second DETECT imputation



An estimate of the Identity by Descent (IBD) for each possible pair of individuals in the dataset was derived in Plink using the option --genome. A pruned set of SNPs with MAF > 0.1 was selected from the cleaned genotyped SNP set (window size=50, number of SNPs to shift after each step=5, variance inflation factor (VIF)=2). In Figure 31, the estimated proportion IBD (PI-HAT) is plotted for the 4 groups of relationships: unrelated individuals (expected PI-HAT 0), full siblings (expected PI-HAT around 0.5), parent-offspring pairs (expectation PI-HAT exactly 0.5), and parent-parent pairs (expected PI-HAT 0). In Figure 32, the estimated proportion IBD 0 is plotted against the estimated proportion IBD 1 for each pair of individuals. The expected proportions IBD 0 – IBD 1 – IBD 2 for several types of family relations are shown in Table 11.

Table 11. Expected proportions IBD 0, IBD 1 and IBD 2 for several type of family relationships.

Relationship	**IBD 0**	**IBD 1**	**IBD 2**
Monozygotic twins	0	0	1
Dizygotic twins / full siblings	.25	5	.25
Parent-offspring	0	.5	0
Avuncular	.5	.5	0
First Cousins	.75	.25	0
Unrelated	1	0	0

Figure 31. Estimated proportion IBD (PI-HAT) for all possible pairs of individuals in the DETECT set, separately for groups expected to be unrelated, full sibling, parent-offspring or parent-parent pairs.

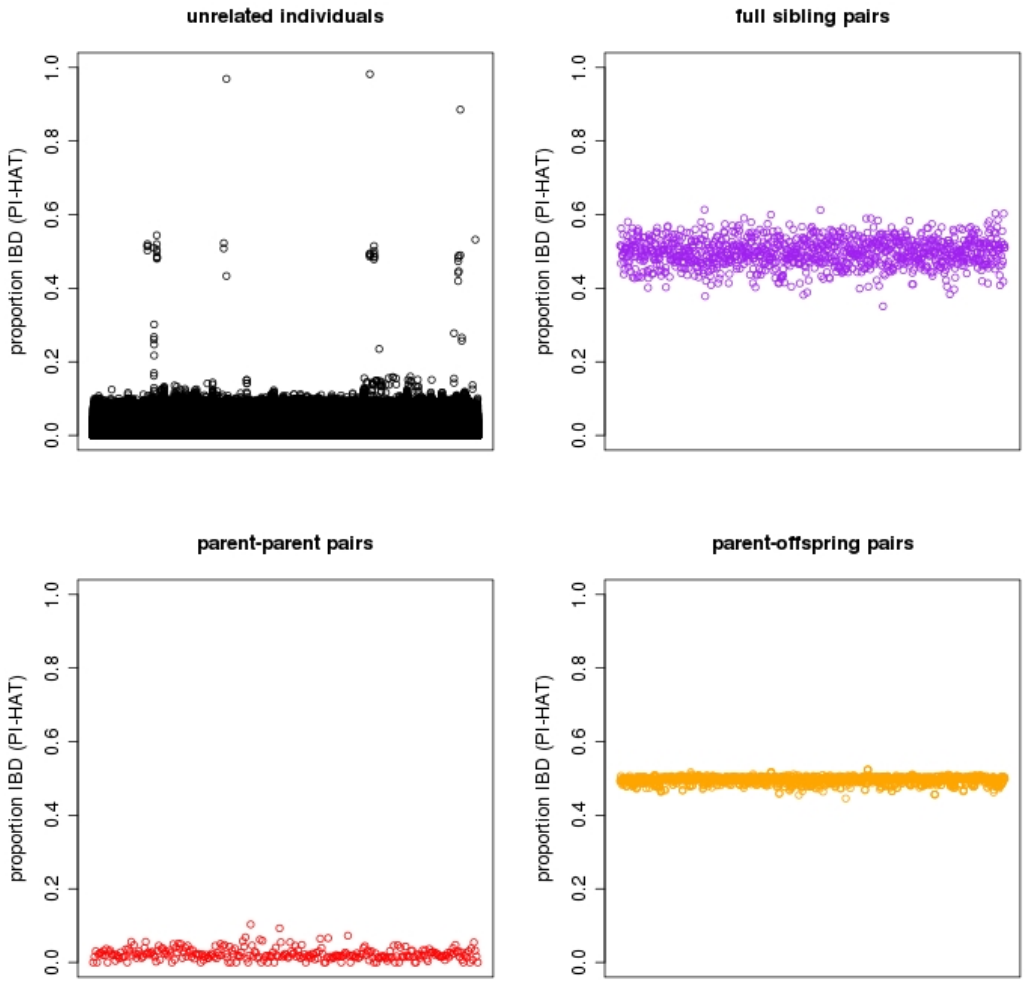
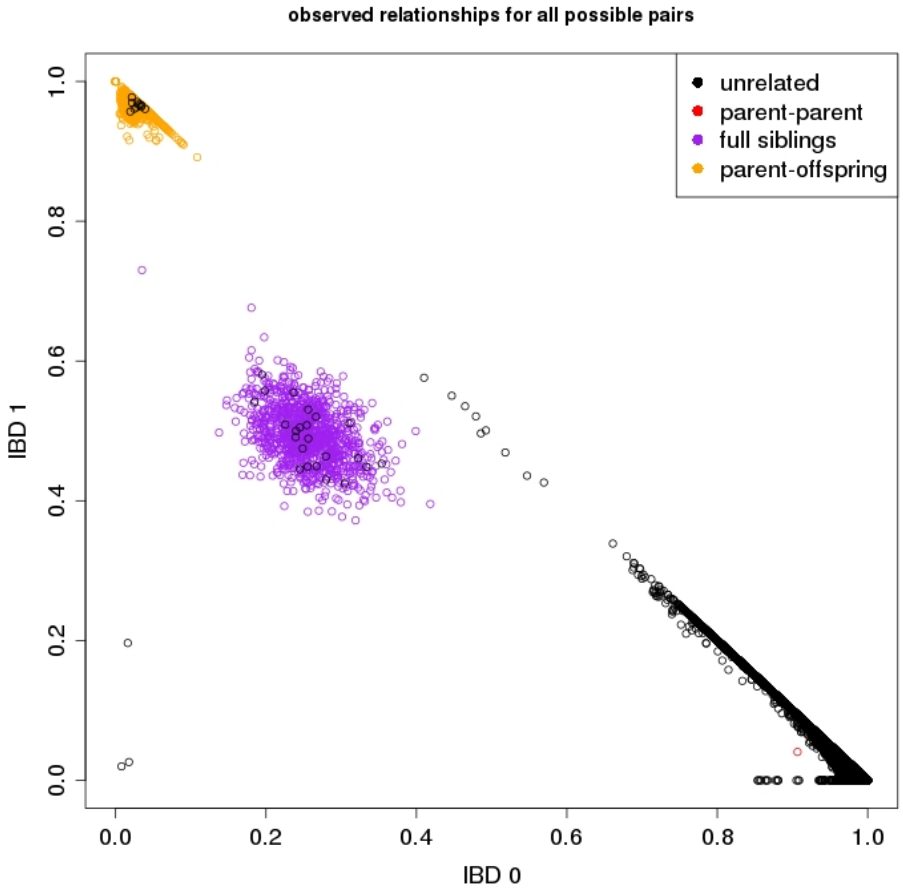


Figure 32. Estimated proportion IBD 0 against the estimated proportion IBD 1 for all possible pairs of individuals in the DETECT set.



A can be seen in the figures, the vast majority of observed relationships was in line with the expectation, but there were some family relationships among individuals thought to be unrelated. These relationships were all checked to see whether the observed relations occurred consistently across all possible family members and whether they were in line with questionnaire information. The family relationships that were judged to be genuine were incorporated in the family structure. For four individuals, the relationships were judged to be improbable and they were excluded from subsequent GWAS analyses. It is interesting to note that the observed relationships for unrelated individuals quite often show a deviation from the expected value of zero. For some pairs these might be genuine, but for the vast majority, these are due to samples with a high heterozygosity, making them appear as genetically similar to all other samples, especially those that also have a higher heterozygosity due to e.g. sample contamination.

Samples were only excluded based on their inbreeding coefficients, not on the above zero relationships with unrelated individuals shown here.

Finally, the quality of the imputation procedure was investigated by repeating the checks for frequency, HWE and Mendelian error rate on the best-guess genotypes in the imputed set. As can be seen in Figure 33-35, the imputation was successful, as, in general, there were no large deviations of the observed frequencies with the frequencies in the reference set, few SNPs with a high number of Mendelian errors and few SNPs with a very small p-value when tested for HWE. As most consortia request analyses results for all SNPs after imputation unfiltered for imputation quality, these SNPs were kept in the dataset.

Figure 33. Frequency of imputed SNPs against the frequency of the same SNP and allele in HapMap release 24 CEU reference panel.

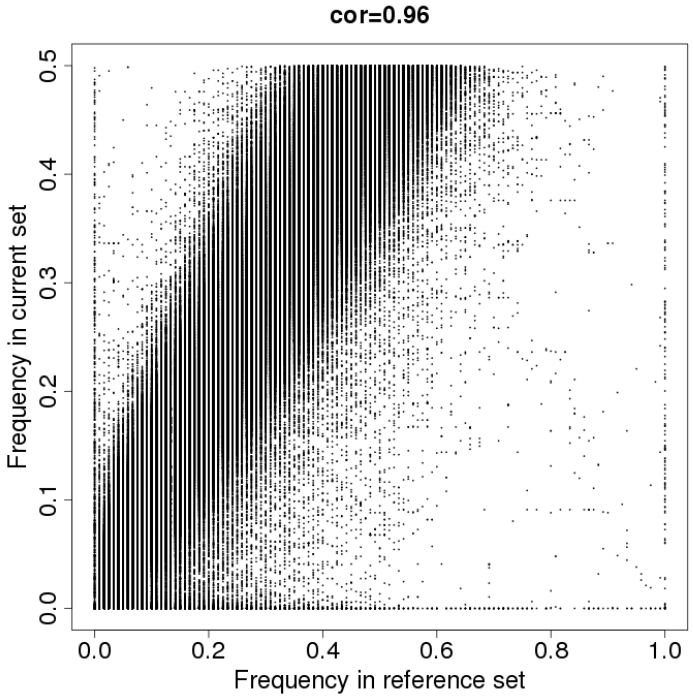


Figure 34. Histogram of the number of Mendelian errors per SNPs for the imputed set (y- axis truncated at 20,000).

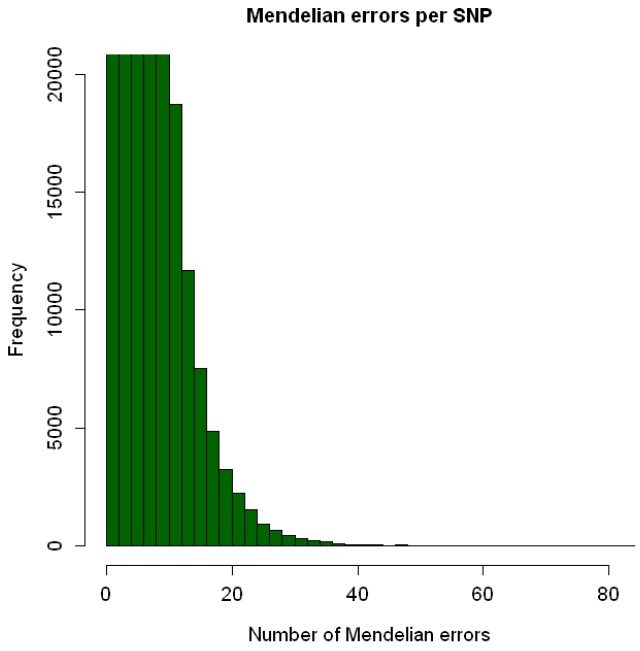
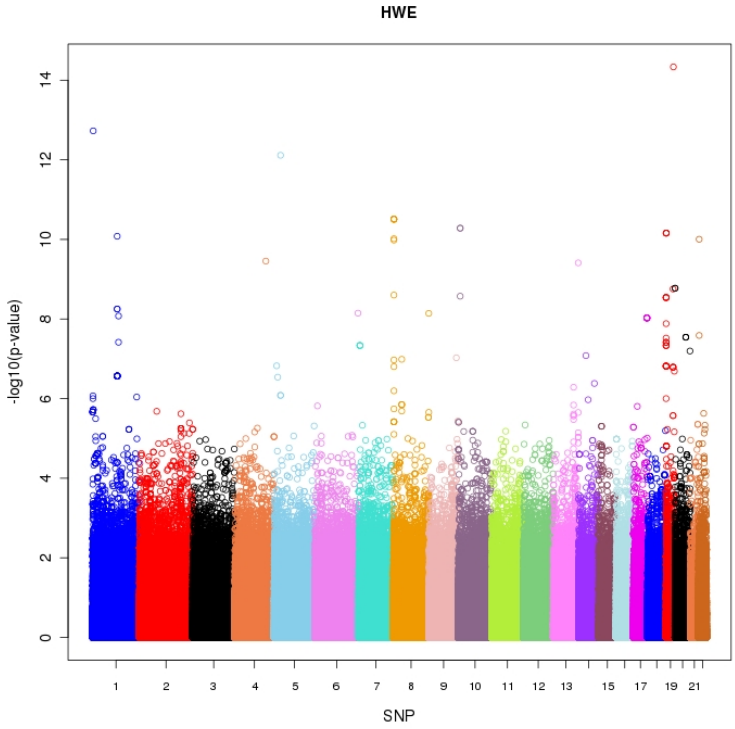


Figure 35. Minus log10 of the p-value of the Hardy-Weinberg equilibrium test for all imputed SNPs.



3. MRG4

Genotyping. In recent years, whole blood and /or buccal DNA samples, were collected for various projects done by the NTR and NESDA studies.[15-18] These sets have been combined and imputed together, as will be described in more detail below. New versions of the total combined set have been created over time, when new samples were added or new reference sets became available. These merged sets are abbreviated as MRG and numbered consecutively. DETECT samples have been included from MRG4 onwards.

DNA extraction and purification of the total set of samples has been performed at various stages in time, following several manufacturer specific protocols in order to obtain the best quality and concentration prior to SNP platform genotyping. The genotyping subsequently has been done on multiple chip platforms, for several partly overlapping subsets of the total sample collection. Chronologically the following platforms have been used Affymetrix Perlegen 5.0 (N=3,840), Illumina 370 (N=290), Illumina 660 (N=1,501), Illumina Omni Express 1M (N=445) and Affymetrix 6.0 (N=8,770, 4 subsets). After array specific data analysis, genotype calls were made with the platform specific software (Genotyper, Beadstudio).

Pre-imputation QC. Quality control has been done within -, as well as between chip platforms. For each platform the individual SNP markers were lifted over to build 37 (HG19) of the Human reference genome, using the LiftOver tool (http://genome.ucsc.edu/cgi-bin/hgLiftOver). SNPs that were not mapped at all, SNPs that had ambiguous locations, and SNPs that did not have matching - or strand opposite alleles were removed. Subsequently, the data were strand aligned with the 1000 Genomes phase I Interim release ALL panel of 23 November 2010 (sequence) and June 2011 (haplotypes) prepared by J. Marchini & B. Howie for IMPUTEv2.1. SNPs from each platform were removed if they still had mismatching alleles with this imputation reference set, if the allele frequencies differed more than 0.20 with the reference set, if the MAF was < 1%, if the HWE p-value was < 0.00001, if the Mendelian error rate was > 2%, if the double typed error rate was > 2% or if the call rate was < 95%. All samples were excluded from the data if their expected sex did not match their genotyped sex, if the genotype missing rate was above 10% or if the Plink measure of elative heterozygsosity F was either > 0.10 or < -0.10. After these steps the data of the individual chips were merged into a single dataset using the Plink 1.07 software.

Within the merged set IBD was calculated between all possible individual pairs and compared to the expected family structure of the NTR and NESDA studies. Samples were removed if the data did not match the expected IBD sharing, or if potentially consistent with biographic data, corrections were made to the family structure. DNA samples that were typed on multiple platforms were tested if the overlapping SNPs had a concordance rate above 99.0%. In case this was not true we removed all data of these samples. On the merged data, the HWE and MAF SNP filters were re-applied, as well as the reference allele frequency difference < 0.20 checks. As a final prior step to imputation SNPs with C/G and A/T allele combinations were removed if the MAF was between 0.35 and 0.50 to avoid wrong strand alignment for these SNPs.

Imputation. Genome wide SNP imputation was done with the IMPUTE 2.1.2 program for the autosomal genome using the above phase I integrated reference panel. A total number of 10,726 unique samples (from MZ twin pairs 1 person was removed) were imputed in batches of around 500 individuals for 5 million base blocks, a 250 kb buffer and the NE parameter set to 20,000. Monomorphic SNPs present in the 500 individual subsets were removed. To avoid issues having SNPs from different platforms partly imputed and partly genotyped we re-imputed all genotyped SNPs. After imputation of these SNPs, we generally find a high concordance between re-imputed SNPs and the original genotype (0.9868), if the SNP info or R2 is above 30%.

Post imputation QC. After imputation, additional QC included evaluation of the SNP platforms effect, the Mendelian error rate in families and filters for HWE, imputation quality and MAF. First we tested the effect of having different platforms imputed, and we removed SNPs showing platform effects. This was done, defining an individual with a specific platform as a case and the remaining individuals as controls. Allelic association was then calculated and SNPs were removed if the specific platform allele frequencies were significantly different from the remaining platforms with $p < 0.00001$. Subsequently, HWE was calculated on the SNPTEST allele probability counts for the full sample and SNPs were removed if the p-value < 0.000001. The Mendelian error rate was calculated on the best guess genotypes in families (trios and sib-pairs with parents) using first Gtool and then Plink 1.07. SNPs were removed if the Mendelian error rate was above 2%. SNPs were subsequently filtered from the dataset if the imputation quality R2 < 0.30 (mean is 0.85) and if the MAF in the total sample was less than 0.004. Finally, all SNPs were removed that were by then known to show problems in the reference sequence

data of the phase I Interim set (159k). As a last step all MZ twins whose zygosity was confirmed by DNA testing were again duplicated in the data. To correct for any remaining confounding by differences in imputation quality across different platforms, chip dummy covariates were created (coded 1 for anyone included on a particular platform and 0 for all others).

4. MRG5

In December 2012, additional GWAS samples were available and another effort was made to combine all available NTR GWAS data, this set comprised 14,003 individuals (including both twins from 1,890 MZ twin pairs). Data were imputed using the 1000 Genomes phase I Integrated release ALL panel of 23 November 2010 (sequence) and 14 March 2012 (haplotypes). Cleaning steps were roughly the same as for MRG4 described above and are not repeated here.

5. GoNL imputation

The same set of individuals that was included in MRG5 was also imputed against the GoNL reference set,[19] but including only one individual of each MZ twin pair, leading to a total of 12,113 individuals.

VII. Principal Component Analysis

Several Principal Component Analyses (PCAs) have been performed by Abdel Abdellaoui in order to identify systematic frequency differences in the NTR genotype datasets. This is an important step in GWA studies because spurious associations can arise when batches or subpopulations differ systematically in both their allele frequencies and their phenotype data. These confounding effects can be (partly) corrected for by including Principal Components (PCs) as covariates in the analyses. In addition, the PCs can be used to identify individuals with non-Dutch / non-European ancestry in order to ensure that the GWAS is performed on a homogenous population. A short description of the PCAs conducted on the NTR data is included here, for a more detailed description, please see Abdellaoui et al., 2013.[20]

To identify individuals with non-Dutch, a PCA was run on genotype data of 1,014 unrelated individuals included in the 1000 Genomes project, coming from 14 different populations across 5 continents. SNPs were selected to overlap with the Affymetrix 6.0 platform and excluded 24 long-range LD regions. These PCs were then projected on the NTR dataset; individuals were identified as non-Dutch / non-European when their PC score was higher or lower than all European individuals in the 1000 Genomes population. 146 Individuals from the DETECT dataset were identified as non-Dutch / non-European and excluded from GWAS analyses.

Subsequently, a PCA was run on a set of 4,666 unrelated individuals genotyped on the Affymetrix 6.0 platform that were of Dutch/European ancestry; this included all DETECT samples, as far as they were unrelated. A pruned SNP set was used excluding 24 long-range LD regions. The first PC from this analysis seemed to reflect a batch effect, caused by a somewhat lower quality of some of the buccal-derived samples. Excluding the samples that were outliers on this PC and using the same procedure, PCs were derived that reflected the Dutch population substructure. The first PC was highly correlated with a north-south gradient and the second PC correlated with an east-west gradient. Finally, a PCA has been ran on the imputed SNPs of MRG4 and MRG5 in order to identify bias due to differential imputation across different chips. A set of unrelated individuals was selected, and 1 million SNPs were selected at random, cleaned, stripped from long-range LD regions and pruned.

VIII. dbGaP upload

As the genotyping of the DETECT samples was funded by the NIMH, it was mandatory that a clean set of genotype and phenotype data was uploaded to the database of Genotypes and Phenotypes (dbGaP).[21] The initial cleaning of the DETECT set led to the exclusion of 60 samples; when the DETECT set was merged with the other NTR genotype dataset an additional 68 samples were selected for exclusion, leaving 3,138 samples for upload. This number included 794 samples of individuals that were part of a monozygotic twin pair (shown by blood or DNA tests), so that genotype data were available for 3,932 individuals. The final upload was done from the VU University by Gonneke Willemsen.

References

1. Paternoster L, Standl M, Chen CM, et al. Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. Nat Genet. 2012;44:187-192.

2. Paternoster L, Standl M, Chen CM, et al. Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. Nat Genet. 2012;44:187-192.

3. van Beijsterveldt CE, Groen-Blokhuis M, Hottenga JJ, et al. The Young Netherlands Twin Register (YNTR): longitudinal twin and family studies in over 70,000 children. Twin Res Hum Genet. 2013;16:252-267.

4. Cousminer DL, Stergiakouli E, Berry DJ, et al. Genome-wide association study of sexual maturation in males and females highlights a role for body mass and menarche loci in male puberty. Hum Mol Genet. Under review.

5. Cousminer DL, Berry DJ, Timpson NJ, et al. Genome-wide association and longitudinal analyses reveal genetic loci linking pubertal height growth, pubertal timing and childhood adiposity. Hum Mol Genet. 2013;22:2735-2747.

6. Perna R, Loughan A. Early developmental delays: neuropsychological sequelae and subsequent diagnoses. Appl Neuropsychol Child. 2012;1:57-62.

7. Pillas D, Hoggart CJ, Evans DM, et al. Genome-wide association study reveals multiple loci associated with primary tooth development during infancy. PLoS Genet. 2010;6:e1000856.

8. Estourgie-van Burk GF, Bartels M, Hoekstra RA, Polderman TJ, Delemarre-van de Waal HA, Boomsma DI. A twin study of cognitive costs of low birth weight and catch-up growth. J Pediatr. 2009;154:29-32.

9. Hoekstra RA, Bartels M, Boomsma DI. Longitudinal genetic study of verbal and nonverbal IQ from early childhood to young adulthood. Learn Individ Differ. 2007;17:97-114.

10. van Dijk BA, Boomsma DI, de Man AJ. Blood group chimerism in human multiple births is not rare. Am J Med Genet. 1996;61:264-268.

11. Derks EM, Hudziak JJ, Dolan CV, Ferdinand RF, Boomsma DI. The relations between DISC-IV DSM diagnoses of ADHD and multi-informant CBCL-AP syndrome scores. Compr Psychiatry. 2006;47:116-122.

12. Nyholt DR. On the probability of dizygotic twins being concordant for two alleles at multiple polymorphic loci. Twin Res Hum Genet. 2006;9:194-197.

13. Sieberts SK, Wijsman EM, Thompson EA. Relationship inference from trios of individuals, in the presence of typing error. Am J Hum Genet. 2002;70:170-180.

14. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet. 2009;84:210-223.

15. Boomsma DI, Willemsen G, Sullivan PF, et al. Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. Eur J Hum Genet. 2008;16:335-342.

16. Scheet P, Ehli EA, Xiao X, et al. Twins, tissue, and time: an assessment of SNPs and CNVs. Twin Res Hum Genet. 2012;15:737-745.

17. Sullivan PF, de Geus EJ, Willemsen G, et al. Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. Mol Psychiatry. 2009;14:359-375.

18. Willemsen G, de Geus EJ, Bartels M, et al. The Netherlands Twin Register biobank: a resource for genetic epidemiological studies. Twin Res Hum Genet. 2010;13:231-245.

19. Boomsma DI, Wijmenga C, Slagboom EP, et al. The Genome of the Netherlands: design, and project goals. Eur J Hum Genet. 2014;22:221-227.

20. Abdellaoui A, Hottenga JJ, de Knijff P, et al. Population structure, migration, and diversifying selection in the Netherlands. Eur J Hum Genet. 2013;21:1277-1285.

21. Mailman MD, Feolo M, Jin Y, et al. The NCBI dbGaP database of genotypes and phenotypes. Nat Genet. 2007;39:1181-1186.

Chapter 8

Summary and General Discussion

The aim of this thesis was to identify genetic and environmental risk factors for behavioral problems, in particular Attention Problems (AP) and Attention Deficit Hyperactivity Disorder (ADHD), and to quantify the contributions of latent and measured genotypes and environmental factors to the observed phenotypic variation. Longitudinal information collected by the Netherlands Twin Register (NTR) over the past 25 years was used to investigate early predictors of AP in childhood. In addition, information from DNA markers available in a subsample of participants was combined with the behavioral data in genetic association studies. Alongside this work, genome-wide association (GWA) studies for other phenotypes in childhood were performed for the EArly Genetics & Lifecourse Epidemiology (EAGLE) and Early Growth Genetics (EGG) consortium; these are two large international consortia of population-based cohorts that perform genetic association studies on a range of childhood phenotypes, including behavioral and somatic traits. Within the EAGLE consortium, I performed a meta-analysis on continuous measures of ADHD symptoms. Data from the Psychiatric Genomics Consortium (PGC),[1] a consortium that focuses on large GWA studies of clinical psychiatric cases and controls including childhood ADHD, were used to test whether the association results based on a study of clinical ADHD were predictive of AP in the NTR. Below I present a summary of the main findings of each chapter included in this thesis; next, I will discuss the results and offer a perspective on future research.

Summary

In **chapter two** I looked at two early indicators of childhood temperament, i.e. items on "crying without a cause" and "being easily upset". These items were included in a questionnaire sent out to parents of 2-year old twins. The heritability of crying without a cause was estimated at 60% in boys and girls; the heritability of being easily upset was estimated at 43% in boys and 31% in girls. The shared environment explained 35 to 63% of the variance. Exploration of the large contribution of the shared environment indicated that it was not accounted for by the influence of birth cohort, gestational age, socioeconomic status, parental age, parental smoking behavior and alcohol use during pregnancy, or personality characteristics of the mother. The association between the two items (polychoric correlation of 0.36) was explained both by genetic and shared environmental factors. Importantly, both items were predictive of internalizing problems,

externalizing problems and Attention Problems 5 years later, at age seven, with effect sizes between 0.28 - 0.42.

Chapter three reports on a study on the effect of birth weight on AP in childhood. In a sample of > 29,000 children, a lower birth weight was associated with higher AP scores at age 3, 7, 10 and 12 years. In monozygotic and dizygotic twins discordant for birth weight, the twin with the lower birth weight scored significantly higher on the continuous AP scale. The differences in AP scores that were observed within unrelated pairs of children and within monozygotic and dizygotic twin pairs were of similar size. This finding strongly supports a causal effect of birth weight on AP. Twin pairs concordant for low birth weight but discordant for catch-up growth showed similar AP scores; the observed relationship between birth weight and AP was therefore unlikely to be explained by catch-up growth in low birth weight children.

Chapter four describes an attempt to replicate a previously reported interaction effect between a measured environmental exposure, namely breastfeeding, and measured genetic polymorphisms of the fatty acid desaturase 2 (FADS2) gene on IQ. These analyses were extended to include overactive behavior at age 3, AP at age 7, 10 and 12, and educational attainment at age 12 as outcome variables in addition to IQ. After correction for maternal education, a small effect of breastfeeding was observed for educational attainment at age 12, overactive behavior at age 3 and IQ across age 5-18; the latter effect was only marginally significant ($p=0.05$). The polymorphisms in the FADS2 gene showed no main effect on any of the included phenotypes; neither did they moderate the effect of breastfeeding.

Chapter five reports on an analysis in which AP in children who take part in NTR research projects were predicted based on polygenic risk scores for ADHD from an independent discovery sample. The results of the most recent ADHD meta-analysis on 5,621 clinical ADHD cases and 13,589 controls, which was carried out by the PGC, were used to calculate polygenic risk scores in Dutch twins and predict their continuous AP scores. Polygenic scores were obtained by multiplying the number of observed effect alleles by the effect size found in the meta-analysis, summed over all loci (a locus refers to a location on the genome where a Single Nucleotide Polymorphism (SNP) was assessed). In the population based NTR sample, the

ADHD polygenic risk scores were predictive of preschool and school-age maternal ratings and school-age teacher ratings of AP.

Chapter six includes the first report of a GWA meta-analysis of continuous measures of AP and ADHD symptoms that was conducted within the EAGLE consortium. Results of nine population-based cohorts were included leading to a total sample size of 17,560 children with genotype data imputed against the 1000 Genomes reference set. The Manhattan plot showed promising signals, but as expected with this sample size, no genome-wide significant findings were found at the stringent p-value threshold of $5E^{-8}$. The tests of individual genetic variants were supplemented with gene-based tests and pathway analyses, but no signals were significant at a false discovery rate of 5%.

Chapter seven documents the results of several more GWA studies of behavioral and somatic phenotypes that were run for the EAGLE and EGG consortia, including atopic dermatitis, childhood and pubertal height, body mass index (BMI) in childhood, pubertal staging and motor development. The NTR results have been uploaded for meta-analysis, together with the results from other childhood cohorts from Australia, Finland, the UK, Germany and other Dutch cohorts like Generation R. Genome-wide significant findings have already been published for eczema and pubertal height growth, and reported from preliminary meta-analyses of pubertal staging and BMI in childhood.[2, 3] This chapter contains a detailed description of the genotyping, data cleaning and imputation procedures that preceded the genetic association analyses.

Discussion

Although the identification of genetic risk factors for psychiatric disorders turns out to be much more involved that initially anticipated, one could argue that an even bigger challenge is to identify the environmental risk factors for psychiatric disorders in children and adults. One design that lends itself for the 'environmental enterprise' is the cotwin control design, which controls for a wide range of confounding factors, thereby eliminating important non-causal explanations for observed associations between environmental exposure and outcome measures such as AP and ADHD. The confounding factors that are controlled for in this design include genetic factors, and environmental factors shared by twins like smoking during pregnancy and gestational age. The NTR data confirmed the association between low birth weight and AP, a finding that was also reported in two previous studies of twins discordant for birth weight. Together, these studies provide strong evidence for a causal relationship between birth weight and ADHD symptoms and underscore the importance of intervention programs that aim to prevent low birth weight.[4, 5] Moreover, they indicate that a close survey of these children at school age may be advisable; AP is negatively correlated with executive functioning, IQ and educational attainment in childhood and there is evidence that low birth weight children with learning disabilities are often not receiving the special academic assistance they need.[6, 7]

A low birth weight is one of the first measureable phenotypes in children. NTR characterizes the early development and environment of twins and multiples by asking mothers about early temperament and breastfeeding. Breastfeeding has been suggested to be beneficial for cognitive development; I looked at the association between breastfeeding on the one hand and IQ, educational attainment and overactive behavior at age 3 on the other hand. In these types of association analyses, maternal IQ and education can confound the association between breastfeeding and these outcomes, as mothers with a higher education are more likely to breastfeed their offspring.[8, 9] Given the high concordance of breastfeeding within twin pairs, it was not possible to apply the cotwin control design to the association of breastfeeding and cognition, but we controlled for maternal education in these analyses, as this information is collected in NTR surveys. The results presented in chapter three and four thus underscore both the importance and difficulty to address issues of confounding.

"No aspect of human behavior genetics has caused more confusion and generated more obscurantism than the analysis and interpretation of the various types of non-additivity and non-independence of gene and environmental action and interaction..." (Eaves et al., 1977).[10] This statement seems as true today as when it was written. I looked at a previously reported gene-environment interaction (GxE) effect between breastfeeding and variants in the FADS2 gene on IQ and did not find evidence for this specific gene-environment interaction; an Australian group earlier reported another non-replication.[11] The only other replication effort performed before found a significant interaction, but in the opposite direction as found in the initial study.[12] These results are representative of the literature on candidate gene association and candidate GxE studies, as the field has been plagued by an inability to detect replicable results. This phenomenon may be due to difficulties to select appropriate candidate genes based on currently limited biological knowledge, and the small effects of individual genetic variants that we have now learned from GWA studies. In the context of small effects, small sample sizes and a publication bias towards positive results; false-positive findings are likely to occur.[13] This issue is further complicated by the fact that interaction effects can be mere statistical phenomena, depending fully on the chosen model and scale of measurement, or anomalies in the data distribution; in these cases non-replication can be due to subtle differences in methodology and data distributions.[14-16] In a critical review of candidate gene GxE findings in psychiatry,[17] many more novel GxE studies than replication attempts turned out to be statistically significant, suggesting that GxE hypotheses appear more robust than they actually are. Given the progress in gene-findings studies since the advance of genome-wide methods, these hypothesis-free approaches should also be applied in GxE research. Finally, another explanation for the lack of robust findings in the gene-environment literature that needs to be considered here, is that gene-environment interactions explain little of the variation in psychopathology. This might not be a fashionable view, but given the fact that little empirical data support the importance of gene-environment interactions so far, this is a possibility one needs to consider seriously.

This thesis includes the first large-scale genome-wide association (GWA) study on continuous measures of Attention Problems and ADHD symptoms based on data in 17,560 children from nine population-based cohorts from the Netherlands, Germany, the UK, Australia, Spain and

Norway. Although this is a large sample size in absolute terms, it is still modest if we compare it to sample sizes of other consortia such as GIANT (the Genetic Investigation of ANthropometric Traits), MAGIC (the Meta-Analyses of Glucose and Insulin-related traits Consortium) and SSGAC (Social Science Genetic Association Consortium) which meta-analyze data in up to 250,000 participants. I carried out a power analysis for AP/ADHD, assuming a normal distribution of the phenotype and no phenotypic heterogeneity. Results showed that we had 80% power to detect a genetic variant that explained 0.23% of the variation (estimated with the Genetic Power Calculator).[18] As we did not detect genetic variants at genome-wide significant levels, it must be assumed that effect sizes are even smaller. The power in our analysis is comparable to the latest PGC meta-analysis of childhood ADHD that included 5,621 clinical ADHD cases and 13,589 controls, and also did not detect genome-wide significant variants.[19] For ADHD, and for other psychiatric phenotypes such as adult major depressive disorder (MDD), it has proven difficult to detect genetic variants in GWA studies. The latest PGC GWA study on major depression found no genome-wide significant hits when including 17,929 cases and 34,693 controls.[20] This is in contrast to the successes that have been reported for schizophrenia in particular, but also bipolar disorder. 108 Loci have been reported in the latest PGC analysis on schizophrenia that included 35,476 cases and 46,839 controls.[21] For bipolar disorder, 8 genome-wide significant hits were reported in a GWA that included 13,741 cases and 19,762 controls.[22] These successes were achieved in samples with much larger statistical power than available for ADHD and MDD; it should be kept in mind that the statistical power depends not only on the sample size, but also on the proportion of cases and the prevalence of the disorder.[23] As the prevalence of schizophrenia and bipolar disorder is much lower than for MDD and ADHD, cases represent a more extreme phenotype, which enhances statistical power substantially. Sample sizes for pediatric samples are generally smaller, and genome-wide studies of SNPs have so far not led to genome-wide significant variants in studies of autism, ADHD, and internalizing behaviors,[19, 24-26] although for e.g. autism findings of the involvement of copy number variants (CNVs) and rare variants have been reported. Similarly, for AP and ADHD, involvement of rare CNVs has been observed.[27, 28]

When analyzing genetic variants that are relatively common, such as SNPs, it is of importance to establish that these genetic variants contribute to variance in ADHD symptoms. This can be done

by using several different methods, including so called chip-based heritability and polygenic risk score analyses. These chip-based heritability analyses employ methods for the estimation of narrow-sense heritability that use measured genome-wide genetic variants in large groups of unrelated subjects, rather than employing the theoretical values of genetic resemblance in relatives. There are two chip-based heritability approaches that differ substantially, with one approach resembling the variance decomposition methods as used in twin studies,[29, 30] and the other based on density estimation (DE) methods.[31] The first method requires raw genotype data and uses these to obtain a measure of genetic similarity between all possible pairs of unrelated individuals in the study. In a second step, this genetic relatedness matrix (GRM) is used to predict the phenotype similarity between individuals. The DE method can be applied after a genome-wide association study has been done. Here, the distribution of z-statistics from a GWAS is compared to the theoretical Null distribution of z-statistics representing no effects. Two studies that used the GRM method to assess the chip-based heritability of ADHD and ADHD symptoms in childhood found contradicting results. Whereas the PGC estimated a chip-based heritability of 28% for the liability to ADHD, a study of self, parent and teacher ratings of ADHD symptoms as measured with the SDQ and Conners' rating scale in the Twins Early Development Study (TEDS) found chip-based heritability estimates close to zero.[32, 33] There is no obvious explanation for this discrepancy, especially in light of the fact that I found that polygenic risk scores based on the PGC ADHD meta-analysis predict AP scores in a population based cohort such as TEDS is as well. It also is important to note that the study from TEDS included a wide range of psychopathology ratings, all with estimates close to zero, in contradiction with other studies that found chip-based heritability estimates of 13-43% for internalizing problems and 18% for social communication traits.[25, 34]

The second approach to test for the relevance of common SNPs makes use of the effect sizes obtained in GWA studies to calculate genetic risk scores based on measured genotypic information in an independent set of individuals. These polygenic scores can then be tested for their ability to predict the phenotype studied in the GWA; if the prediction is significant, this indicates that the effect estimates from the GWA in aggregate contain a relevant signal for the disease or trait. With regard to ADHD, ADHD polygenic risk scores have been shown to be predictive of ADHD case status, and, as described in chapter five of this thesis, of measures of

Attention Problems in a population-based samples.[35, 36] The latter hints to the possibility to use polygenic scores to test for the genetic overlap across traits, in which polygenic risk scores based on one trait are used to predict the other trait in an independent sample. Such a cross-trait analysis is also possible with the GRM method, as it allows for the estimation of a chip-based co-heritability of two traits. Thereby, these methods contribute to the exciting shift from the analysis of single traits to cross-disorder studies. The PGC group, for example, has performed a GWA of 33,332 cases of five psychiatric disorders (autism spectrum disorder, attention deficit hyperactivity disorder, bipolar disorder, major depressive disorder, and schizophrenia) and 27,888 controls and detected four loci that contribute to variation across these disorders.[37] Moreover, using both polygenic risk score and chip-based co-heritability methods, they found convincing evidence for a genetic overlap across several psychiatric disorders, most convincingly bipolar disorder and schizophrenia.[32, 37] These studies thus show that even when individual genetic variants confer only small relative risks, analyses that consider all risk variants together can contribute to an enhanced understanding of the etiology of psychiatric traits. In addition, they can be used to investigate the heterogeneity of psychiatric disorders and the genetic associations with endophenotypes, such as cognitive functioning.

Before discussing how research on AP and ADHD should proceed, I will also discuss the findings of this thesis in the light of the ongoing debate, both in the literature and in the popular media, on the validity of ADHD as a clinical psychiatric disorder. Opponents claim that ADHD is a social construct leading to unnecessary medicalization of difficult children. They argue that there is no clear physical cause of ADHD and that the steep increase in ADHD diagnoses and stimulant use over time imply an important influence of social factors. Others defend the ADHD concept by pointing out that the observed behavior constitutes an impairment to the child and is associated with long-term negative consequences.[38] Although large heritability estimates and reported genetic associations are sometimes taken as evidence for the reality of ADHD as a disorder, it is important to note that the question whether ADHD is a 'real' medical condition or a social construct cannot be answered by only looking at genetic effects. Many so-called social constructs such as education or socioeconomic status turn out to be influenced by genes, while some somatic diseases such as lung cancers clearly have a huge non-genetic component to explain their etiology.

Yet, genetic analyses are of importance to aid our understanding of ADHD. The polygenic risk score analysis described in chapter five indicates that the genetic liability for ADHD is continuously distributed. This implies that the clinical cutoff applied in clinical practice is in essence arbitrary; one's score on the ADHD scale can be more or less extreme, with the subsequent impairment also partly depending on the context in which you are functioning. This idea is further supported by the fact that in the NTR the number of reported AP and ADHD symptoms has stayed constant over the last two decades, while the number of ADHD diagnoses and stimulant prescriptions has seen a steep increase in this period, as indeed also reported by parents of NTR twins, thereby signifying that which score on the continuous scale is considered impairing is partly influenced by social factors. Moreover, it is of interest to note that children in the youngest one third of a class have a 50% increased risk of being prescribed stimulant medication when compared to the oldest one third. This is probably due to the fact that the same behavior is considered more impairing when children are surrounded by somewhat older children that have matured to higher levels of self-control.[39]

The notion that children with a certain level of ADHD symptoms are now more often considered clinical cases than before, does not answer the question whether this is because the clinical disorder is now recognized more often, or whether the increase is due to misclassification, or whether the number of symptoms stayed the same while the environment of the child became less tolerant of the behaviors. In the end, questions on the validity of the ADHD diagnosis can only be answered by a careful reflection on the nature of psychiatric disorders, the borders between psychiatry and normality and the difference between treatment and enhancement; a clear-cut answer cannot be provided. Given the earlier mentioned negative long-term consequences and impairment associated with extreme scores in current society, there is no reason to put ADHD aside as a nonsense diagnosis; the impairment is a reality that needs to be reacted upon. Yet, the understanding of ADHD as the extreme end of a continuous distribution can hopefully help to lessen stigmatization and to keep an open view on the possibilities of treatment for individuals at threshold and subthreshold levels of the disorder.

Future prospects

If we accept that there is convincing evidence for the implication of genetic variants in the etiology of psychiatric traits and ADHD, it is clear that GWA studies will continue to play an important role in genetic studies of ADHD in the near future. Such studies will enlighten us about the genetic architecture of traits and will, maybe paradoxically, also aid research on environmental factors by enabling methods that go beyond establishing an association between the environment and a trait, to explain the underlying cause of the association.

An important question is how GWA studies should move forward in order to find replicable associations. Clearly, increasing sample sizes is of major importance for success; current efforts to collect GWA data in larger sets of ADHD cases are thus worthwhile. Several issues need to be considered when one aims to increase sample sizes for GWA studies. It is for example, of importance to look at the implications of the organizational structure of GWA consortia. Two different strategies are currently followed; the PGC is the most prominent example of a consortium that requires all participating cohorts to upload their individual-level genotype and phenotype data to a common server. The main advantage of this approach is that quality control, imputation and analyses can be performed on the total dataset, enhancing uniformity across cohorts. Once the data have been uploaded, the speed of the project is dependent on the consortium analysts only; in general, this will speed up the process. Moreover, this approach allows for more complex analyses for which individual level data are required, such as the GRM chip-based heritability estimates. A clear downside of this approach is that the requirement of individual level data sharing hampers the number of cohorts that are able or willing to participate. Sharing of individual level data comes with many ethical and legal issues, so that some cohorts will not be able to contribute. Moreover, it is not easy to find a balanced system to distribute academic credit over the contributors; once data have been uploaded, contributing cohorts are in a somewhat difficult position to negotiate. To bypass these problems, other consortia only require the sharing of GWA summary statistics, an approach that is taken in for example ENIGMA, MAGIC and GIANT and also in the EAGLE and EGG consortia. To be effective, it is essential that there is a clear and comprehensive analysis plan and that cohorts are able to provide high quality results within a reasonable time frame. In practice, this strategy has proven to work well, both in the EAGLE and EGG consortia as in other consortia of even larger

scale, such as the SSGAC that recently included 126,559 individuals from 54 cohorts in a report on three genome-wide significant SNPs for educational attainment.[40]

Another important issue when it comes to increasing sample size is the issue of phenotype definition. For studies focusing on gene finding rather than precise effect estimation, a cost-effective approach may be to collect large samples with less in-depth, but easier to obtain phenotypic information than extensive clinical assessments; this type of phenotype measure is often collected in large twin registers. The disadvantage of survey measures may be compensated by the increase in sample size and availability of longitudinal phenotype information; using this information may be a promising approach for future projects. Another option is the use of electronic medical records and information on medication use as a proxy for case status, or the use of single questionnaire items that ask a participant whether he has ever been diagnosed with a particular disease. The latter can be a fruitful approach, as shown by the GWA on atopic dermatitis reported in chapter seven, and studies that successfully replicated known genetic associations with self-reported health data.[3, 41] However, it should be noted that the use of a single item phenotype definition will probably be more challenging for diseases with a low prevalence, such as for ADHD (the prevalence of ADHD is around 5%, whereas eczema has a prevalence around 15-30% in childhood).[42] In these cases, the collection of behavior questionnaires that assess the full range of behaviors in the population remains an attractive alternative, as was done in the GWA meta-analysis on continuous measures of ADHD symptoms described in chapter six. These population-based cohorts have further advantages in terms of cost-effectiveness, as they generally assess a wide range of phenotypes, instead of being focused on one particular disease only. Another worthwhile option is to perform a meta-analysis in which data from clinical case-control and population-based cohorts are analyzed together. Although the heterogeneity in phenotypic assessment will lead to a decrease in statistical power, this effect could very well be outweighed by the increase in sample size. With regard to ADHD, it will be worthwhile to start with a combined meta-analysis of the PGC and EAGLE data. As an extension of such a cross-consortium meta-analysis, the inclusion of adult data could be considered, as behavior genetic studies have found a considerable overlap of genetic variants for ADHD symptoms across the life span.[43]

Obviously, this approach is only fruitful if the genetic variants for a clinical diagnosis of ADHD overlap to a sufficient extent with genetic variants that influence ADHD symptoms in the general population. In line with previous behavior genetic studies based on twin data, our polygenic risk score analysis indicated a significant genetic overlap between the genetic variants detected by GWA of clinical ADHD cases and continuous measures of ADHD as measured with the CBCL AP scale in the NTR in childhood.[44, 45] As a follow-up on this analysis, I tested for the overlap between the GWA meta-analysis results of PGC and EAGLE; the latter included a considerably larger number of children and a wider variety of measures of ADHD symptoms in terms of measurement instrument, age and rater. Selecting independent top SNPs from the PGC analyses after appropriate quality control, the top hits of PGC did not show significant overall inflation of p-values in EAGLE (as indicated by the inflation factor lambda). However, the direction of effect was significantly more consistent across the consortia than expected by chance, as indicated by a sign-test (see Table 1), with a stronger level of significance at less stringent p-value thresholds of inclusion. This consistency of effect direction at more liberal p-value thresholds shows that there is a relevant genetic signal in the PGC ADHD mega-analysis that overlaps with the genetic signal in the EAGLE meta-analysis of ADHD symptoms, providing support for the above-mentioned possibility to perform a cross-consortium meta-analysis of PGC and EAGLE. At the same time, the most significant top hits from PGC (p-value < $1E^{-5}$) showed a similar direction of effect in the EAGLE meta-analysis in only 50% of the cases; this again underscores the difficulty to identify individual genetic variants for ADHD with current sample sizes. It will therefore be worthwhile to not focus too extensively on the top hits, but instead search for replication for a longer list of variants, thereby incorporating also biological information in the process of prioritizing variants for follow-up analyses.

Table 1. Overall inflation and consistency of direction of effect in the EAGLE ADHD symptoms meta-analysis for the independent top SNPs from the PGC ADHD case-control mega-analysis.

SNP set, p-value threshold PGC results	n SNPs	Lambda	empirical p-value	Sign-test (same/different direction)	p-value sign-test
$1E^{-5}$	12	1.813	0.1555	6/6	1
$1E^{-4}$	75	0.808	0.7808	45/30	0.1053
$1E^{-3}$	600	0.965	0.6621	322/278	0.7909
$1E^{-2}$	4,563	0.992	0.6526	2,409/2,150	$1.3E^{-4}$
$1E^{-1}$	32,623	1.021	0.0766	17,045/15,562	$2.2E^{-16}$
1	168,871	1.005	NA	85,735/82,747	$3.4E^{-13}$

Different methods have been developed to analyze the genetic architecture of traits and their chip-based heritability. These methods have not yet been extensively tested for their robustness against violations of assumptions. Moreover, the effects of specific thresholds applied in quality control steps are often unknown. It is currently, for example, unclear how quality control filtering affects the estimates of explained variance from the GRM and DE methods, and the outcomes of polygenic risk score analyses. This includes filtering on MAF, imputation quality and HWE, but also correction for patterns of linkage disequilibrium. The latter is an important practical issue, obtaining a set of independent variants is essential for e.g. the DE method, and may be of importance for polygenic risk score analyses; but both the appropriate methods and cutoffs are yet to be agreed on. In studies that apply polygenic risk score analyses, results are often reported for both pruned and un-pruned datasets, but whenever results are dissimilar, it is hard to judge which result is more appropriate. For the GRM method, several options have been proposed to correct for LD, but it remains to be established which method performs best under which scenario.[46, 47]

With regard to GWA studies, outstanding issues include variability in QC thresholds both at the individual cohort and meta-analysis level, the appropriate cutoff for genome-wide significance when analyzing 1000 Genomes imputed datasets and the use of fixed versus random effect models. Obviously, this situation is a consequence of the pace of development in the field, and is as such inescapable. Nevertheless, studies that look further into these issues are timely and

needed. Where discrepancies have been found such as for the above mentioned differences in chip based heritability estimates of ADHD and ADHD symptom scores, these studies may help to elucidate whether these differences can be attributed to methodological differences. The last year has seen a positive development in this regard. Papers by Dudbridge et al. and Lee et al. on polygenic risk score analyses provided a helpful examination of the behavior of these tests under different scenarios, although they do not address the issue of pruning.[48, 49] With regard to the GRM method, a recent paper has tested the effects of violations of five main assumptions.[47] More of these methodological issues will likely consolidate in the near future, increasing the validity and consistency in application of these methods.

In the meanwhile, research on environmental risk factors should also proceed. Several methods have been suggested to test for causality of associations with environmental risk factors, including randomized controlled trials, statistical control for measured confounding factors, twin and family designs and Mendelian randomization.[50] While each method has its own limitations, together these methods can provide convincing support for causal effects of established risk factors. So far, they have been applied somewhat sparsely to environmental risk factors for ADHD. In addition to the two studies, I have performed, two other studies, one applying a discordant sibling design and another employing an assisted reproduction designs of mothers that are genetically unrelated to their child, suggest that the relationship between smoking during pregnancy with ADHD is at least partly due to confounding by genetic and other familial factors.[51, 52] A sibling study found that the association between low income during early childhood and ADHD remained after control for familial factors.[53] Further use of these methods can hopefully contribute to the clarification of the role of the environment in ADHD and identify genuine risk factors.

The lack of significant findings in GWA studies is sometimes attributed to the hypothesized contribution of gene-environment or gene-gene interactions to the phenotypic variation. However, given the above-mentioned lack of empirical evidence for the contribution of gene-environment interaction to psychopathology, the initial focus of current studies should be on the identification of robust genetic associations. Subsequently, gene-environment or gene-gene interactions can be investigated. Given the growing number of established associations and the

clear predictive value of polygenic scores, studies that combine GWA data with measured environmental factors are a clear way forward to the field.[54, 55] One example of such a study is provided in chapter seven, where we contributed to a genome-wide study on the interaction of smoking during pregnancy and genetic variants on birth weight. A similar argument can be made for studies on gene-gene interactions. Although the number of possible interactions is extremely large and the power for interaction studies is generally lower than for main effects, the multiple testing burden can be reduced substantially by including only interactions between known loci, a strategy that was used in the EAGLE GWA on atopic dermatitis, reported in chapter seven.

It is not expected that genetic variants will be suitable for risk prediction and prevention in the near future.[56] However, there have already been examples of pharmacological agents based on genetic findings.[57] Such examples show that the identification of genetic variants can be very valuable, even when they do not explain all heritability. In the end, the hope of many who work in psychiatric genetics may thus prove grounded; that psychiatric genetic studies can make an important contribution to the improvement of mental health.

References

1. Sullivan PF. The psychiatric GWAS consortium: big science comes to psychiatry. Neuron. 2010;68:182-186.

2. Cousminer DL, Berry DJ, Timpson NJ, et al. Genome-wide association and longitudinal analyses reveal genetic loci linking pubertal height growth, pubertal timing and childhood adiposity. Hum Mol Genet. 2013;22:2735-2747.

3. Paternoster L, Standl M, Chen CM, et al. Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. Nat Genet. 2012;44:187-192.

4. Asbury K, Dunn JF, Plomin R. Birthweight-discordance and differences in early parenting relate to monozygotic twin differences in behaviour problems and academic achievement at age 7. Dev Sci. 2006;9:F22-F31.

5. Hultman CM, Torrang A, Tuvblad C, Cnattingius S, Larsson JO, Lichtenstein P. Birth weight and attention-deficit/hyperactivity symptoms in childhood and early adolescence: a prospective Swedish twin study. J Am Acad Child Adolesc Psychiatry. 2007;46:370-377.

6. Litt J, Taylor HG, Klein N, Hack M. Learning disabilities in children with very low birthweight: prevalence, neuropsychological correlates, and educational interventions. J Learn Disabil. 2005;38:130-141.

7. Polderman TJ, Gosso MF, Posthuma D, et al. A longitudinal twin study on IQ, executive functioning, and attention problems during childhood and early adolescence. Acta Neurol Belg. 2006;106:191-207.

8. Bartels M, van Beijsterveldt CE, Boomsma DI. Breastfeeding, maternal education and cognitive function: a prospective study in twins. Behav Genet. 2009;39:616-622.

9. Der G, Batty GD, Deary IJ. Effect of breast feeding on intelligence in children: prospective study, sibling pairs analysis, and meta-analysis. BMJ. 2006;333:945.

10. Eaves LJ, Last K, Martin NG, Jinks JL. Progressive Approach to Non-Additivity and Genotype-Environmental Covariance in Analysis of Human Differences. Brit J Math Stat Psy. 1977;30:1-42.

11. Martin NW, Benyamin B, Hansell NK, et al. Cognitive function in adolescence: testing for interactions between breast-feeding and FADS2 polymorphisms. J Am Acad Child Adolesc Psychiatry. 2011;50:55-62 e54.

12. Steer CD, Davey Smith G, Emmett PM, Hibbeln JR, Golding J. FADS2 polymorphisms modify the effect of breastfeeding on child IQ. PLoS One. 2010;5:e11570.

13. Ioannidis JP. Why most published research findings are false. PLoS medicine. 2005;2:e124.

14. Eaves LJ. Genotype x Environment interaction in psychopathology: fact or artifact? Twin Res Hum Genet. 2006;9:1-8.

15. Kendler KS, Gardner CO. Interpretation of interactions: guide for the perplexed. Br J Psychiatry. 2010;197:170-171.

16. Zammit S, Owen MJ, Lewis G. Misconceptions about gene-environment interactions in psychiatry. Evid Based Ment Health. 2010;13:65-68.

17. Duncan LE, Keller MC. A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. Am J Psychiatry.168:1041-1049.

18. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics. 2003;19:149-150.

19. Holmans P. Pathway analysis of large multicentre ADHD genome-wide datasets. Abstract of oral presentation 21st World Congress of Psychiatric Genetics; October 17-21, 2013; Boston, USA.

20. Breen G, Psychiatric Genomics Consortium Major Depressive Disorder Working Group. The PGC2 for major depressive disorder. Abstract of oral presentation 21st World Congress of Psychiatric Genetics; October 17-21, 2013; Boston, USA.

21. Ripke S, Psychiatric Genomics Consortium Schizophrenia Working Group. Psychiatric genomics Consortium quadruples schizophrenia GWAS sample-size to 35,000 cases and 47,000 controls. Abstract of oral presentation 21st World Congress of Psychiatric Genetics; October 17-21, 2013; Boston, USA.

22. Charney AW, Ruderfer D, Moran J, et al. Bipolar disorder GWAS of 13,741 cases and 19,762 controls identifies eight genome-wide significant hits and implicates genes targeted by FMRP protein. Abstract of oral presentation 21st World Congress of Psychiatric Genetics; October 17-21, 2013; Boston, USA.

23. Yang J, Wray NR, Visscher PM. Comparing apples and oranges: equating the power of case-control and quantitative trait association studies. Genet Epidemiol. 2010;34:254-257.

24. Anney RJL, Ripke S, Arking D, et al. Meta-analysis of European ancestry individuals with autism spectrum disorder reveals strong association 3' of the astrotactin 2 (ASTN2) gene locus on chromosome 9. Paper Abstract of oral presentation 21st World Congress of Psychiatric Genetics2013; Boston, USA.

25. Benke KS, Nivard MG, Velders FP, et al. A genome-wide association meta-analysis of preschool internalizing problems. J Am Acad Child Adolesc Psychiatry. Accepted for publication.

26. Trzaskowski M, Eley TC, Davis OS, et al. First Genome-Wide Association Study on Anxiety-Related Behaviours in Childhood. PLoS ONE. 2013;8:e58676.

27. Ehli EA, Abdellaoui A, Hu Y, et al. De novo and inherited CNVs in MZ twin pairs selected for discordance and concordance on Attention Problems. Eur J Hum Genet. 2012;20:1037-1043.

28. Williams NM, Zaharieva I, Martin A, et al. Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. Lancet. 2010;376:1401-1408.

29. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76-82.

30. Yang J, Manolio TA, Pasquale LR, et al. Genome partitioning of genetic variation for complex traits using common SNPs. Nat Genet. 2011;43:519-525.

31. So HC, Li M, Sham PC. Uncovering the total heritability explained by all true susceptibility variants in a genome-wide association study. Genet Epidemiol. 2011;35:447-456.

32. Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee SH, Ripke S, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet. 2013;45:984-994.

33. Trzaskowski M, Dale PS, Plomin R. No genetic influence for childhood behavior problems from DNA analysis. J Am Acad Child Adolesc Psychiatry. 2013;52:1048-1056.e3.

34. St Pourcain B, Whitehouse AJ, Ang WQ, et al. Common variation contributes to the genetic architecture of social communication traits. Mol Autism. 2013;4:34.

35. Groen-Blokhuis MM, Middeldorp CM, Abdellaoui A, et al. Attention Deficit Hyperactivity Disorder polygenic risk scores predict Attention Problems in a population-based sample of children. J Am Acad Child Adolesc Psychiatry. Under review.

36. Hamshere ML, Langley K, Martin J, et al. High loading of polygenic risk for ADHD in children with comorbid aggression. Am J Psychiatry. 2013;170:909-916.

37. Cross-Disorder Group of the Psychiatric Genomics C, Genetic Risk Outcome of Psychosis C. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet. 2013;381:1371-1379.

38. Willcutt EG, Carlson CL. The diagnostic validity of attention-deficit/hyperactivity disorder. Clin Neurosci Res. 2005;5:219-232.

39. Zoega H, Valdimarsdottir UA, Hernandez-Diaz S. Age, academic performance, and stimulant prescribing for ADHD: a nationwide cohort study. Pediatrics. 2012;130:1012-1018.

40. Rietveld CA, Medland SE, Derringer J, et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science. 2013;340:1467-1471.

41. Tung JY, Do CB, Hinds DA, et al. Efficient replication of over 180 genetic associations with self-reported medical data. PLoS ONE. 2011;6:e23473.

42. Bieber T. Atopic dermatitis. N Engl J Med. 2008;358:1483-1494.

43. Kan K, Dolan CV, Nivard MG, et al. Genetic and environmental stability in attention problems across the lifespan: evidence from the Netherlands twin register. J Am Acad Child Adolesc Psychiatry. 2013;52:12-25.

44. Larsson H, Anckarsater H, Rastam M, Chang Z, Lichtenstein P. Childhood attention-deficit hyperactivity disorder as an extreme of a continuous trait: a quantitative genetic study of 8,500 twin pairs. J Child Psychol Psychiatry. 2012;53:73-80.

45. Levy F, Hay DA, McStephen M, Wood C, Waldman I. Attention-deficit hyperactivity disorder: a category or a continuum? Genetic analysis of a large-scale twin study. J Am Acad Child Adolesc Psychiatry. 1997;36:737-744.

46. Lee SH, Yang J, Chen GB, et al. Estimation of SNP heritability from dense genotype data. Am J Hum Genet. 2013;93:1151-1155.

47. Speed D, Hemani G, Johnson MR, Balding DJ. Improved heritability estimation from genome-wide SNPs. Am J Hum Genet. 2012;91:1011-1021.

48. Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 2013;9:e1003348.

49. Lee SH, Wray NR. Novel genetic analysis for case-control genome-wide association studies: quantification of power and genomic prediction accuracy. PLoS One. 2013;8:e71494.

50. Lewis SJ, Relton C, Zammit S, Smith GD. Approaches for strengthening causal inference regarding prenatal risk factors for childhood behavioural and psychiatric disorders. J Child Psychol Psychiatry. 2013;54:1095-1108.

51. D'Onofrio BM, Van Hulle CA, Waldman ID, et al. Smoking during pregnancy and offspring externalizing problems: an exploration of genetic and environmental confounds. Dev Psychopathol. 2008;20:139-164.

52. Thapar A, Rice F, Hay D, et al. Prenatal smoking might not cause attention-deficit/hyperactivity disorder: evidence from a novel design. Biol Psychiatry. 2009;66:722-727.

53. Larsson H, Sariaslan A, Langstrom N, D'Onofrio B, Lichtenstein P. Family income in early childhood and subsequent attention deficit/hyperactivity disorder: a quasi-experimental study. J Child Psychol Psychiatry. 2013.

54. McGrath JJ, Mortensen PB, Visscher PM, Wray NR. Where GWAS and epidemiology meet: opportunities for the simultaneous study of genetic and environmental risk factors in schizophrenia. Schizophr Bull. 2013;39:955-959.

55. Thomas D. Methods for investigating gene-environment interactions in candidate pathway and genome-wide association studies. Annu Rev Public Health. 2010;31:21-36.

56. van Dongen J, Slagboom PE, Draisma HH, Martin NG, Boomsma DI. The continuing value of twin studies in the omics era. Nat Rev Genet. 2012;13:640-653.

57. Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. Nature reviews. Drug discovery. 2013;12:581-594.

Acknowledgements

I am very grateful for the grants that made the work described in this thesis possible. The European Research Council Advanced Grant that was awarded to Dorret Boomsma (Genetics of Mental Illness, ERC-230374) provided the financial support for my PhD trajectory at the VU University. The ARRA RC2 2MH08995, granted by the National Institute of Mental Health, financed a large genotyping project of samples from the Young Netherlands Twin Register, thereby making an essential contribution to the genetic analyses presented here. Finally, I would like to thank the EMGO+ institution for providing the overall research framework in which this project was embedded.