Genome-wide association scan of the time to onset of Attention Deficit Hyperactivity Disorder

Jessica Lasky-Su[5], **Richard J.L. Anney**[6], **Benjamin M. Neale**[1,2,3,4], **Barbara Franke**[7,8], **Kaixin Zhou**[1], **Julian B. Maller**[9], **Alejandro Arias Vasquez**[7,8], **Wai Chen**[1], **Philip Asherson**[1], **Jan Buitelaar**[7], **Tobias Banaschewski**[10], **Richard Ebstein**[11], **Michael Gill**[6], **Ana Miranda**[12], **Fernando Mulas**[22], **Robert D. Oades**[13], **Herbert Roeyers**[14], **Aribert Rothenberger**[15], **Joseph Sergeant**[16], **Edmund Sonuga-Barke**[1,18,19], **Hans Christoph Steinhausen**[17], **Eric Taylor**[1], **Mark Daly**[3,4], **Nan Laird**[20], **Christoph Lange**[5,20], and **Stephen V. Faraone**[2,21]

[1]Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, UK [2]Department of Psychiatry, SUNY Upstate Medical University, Syracuse, New York, USA [3]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA [4]The Broad Institute of Harvard and MIT, Cambridge Massachusetts, USA [5]Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA [6]Department of Psychiatry, Trinity Centre for Health Sciences, St. James's Hospital, Dublin, Ireland [7]Department of Psychiatry, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands [8]Department of Human Genetics, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands [9]Wellcome Trust Centre for Human Genetics, Oxford University, Oxford, United Kingdom [10]Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute Mental Health, University of Heidelberg, Mannheim, Germany [11]Geha MHC, Petach-Tikva, Israel [12]Department of Developmental and Educational Psychology, University of Valencia, Spain [13]University Clinic for Child and Adolescent Psychiatry, Essen, Germany [14]Ghent University, Dunantlaan, Ghent, Belgium [15]Child and Adolescent Psychiatry, University of Gottingen, Gottingen, Germany [16]Vrije Universiteit, De Boelelaan, Amsterdam, The Netherlands [17]Department of Child and Adolescent Psychiatry, University of Zurich, Switzerland [18]School of Psychology, Institute for Disorder on Impulse and Attention, University of Southampton, Highfield, Southampton SO17 1BJ, UK [19]Child Study Center, New York University, New York, NY 10016, USA [20]Harvard School of Public Health, Boston, MA 02115, USA [21]Department of Neuroscience, SUNY Upstate Medical University, Syracuse, New York, USA [22]Department of Neuropaediatrics, La Fe University Hospital, Valencia, Spain

Abstract

A time-to-onset analysis for family-based samples was performed on the genomewide association (GWAS) data for Attention Deficit Hyperactivity Disorder (ADHD) to determine if associations exist with the age at onset of ADHD. The initial dataset consisted of 958 parent-offspring trios that were genotyped on the Perlegen 600,000 SNP array. After data cleaning procedures, 429,981 autosomal SNPs and 930 parent-offspring trios were used found suitable for use and a family-based logrank analysis was performed using that age at first ADHD symptoms as the quantitative trait of interest. No SNP achieved genome-wide significance, and the lowest p-values had a magnitude of 10^{-7}. Several SNPs among a pre-specified list of candidate genes had nominal associations including SLC9A9, DRD1, ADRB2, SLC6A3, NFIL3, ADRB1, SYT1, HTR2A,

Corresponding author: Dr. Stephen Faraone, Department of Psychiatry and Behavioral Sciences, SUNY Upstate Medical University, 750 East Adams St., Syracuse, NY 13210; faraones@upstate.edu, 315-464-3113, (315) 464-3255 (fax).

ARRB2, and CHRNA4. Of these findings SLC9A9 stood out as a promising candidate, with nominally significant SNPs in six distinct regions of the gene.

SUMMARY

Attention deficit hyperactivity disorder (ADHD) (AHC [MIM 143465]) is a complex childhood onset disorder of inattention, hyperactivity and impulsivity (American Psychiatric Association 1994). Age at onset has often been used as an informative phenotype for studying the genetic basis of psychiatric disorders (Chen and others 1992; Faraone and others 2004); however, ADHD individuals have seldom been used in time-to-onset analyses, primarily because the onset of the disease occurs early in life. The narrow age range of this phenotype makes it more difficult to separate individuals into distinct groupings on the basis of their genotypic information. Lasky-Su(2007) used the age at onset of ADHD as a phenotype to study the association between five candidate genes for ADHD and found that SNPs surrounding the D5 dopamine receptor (*DRD5*) gene were associated with the age at onset of ADHD. Interestingly, this study also found that a large number of individuals with the risk genotype indicated that their ADHD began in infancy whereas the individuals without the risk genotype generally exhibited ADHD behaviors in early childhood. This study suggested that using the age at onset of ADHD as a phenotype can be informative for genetic association studies and that environmental exposures outside of the womb were less likely to be important contributors to the development of the very early onset form of the disorder.

This prior work suggests that using age at onset as a phenotype in genomewide association studies (GWAS) for identifying ADHD susceptibility genes could be useful. In this manuscript, we use the age at onset of ADHD in a logrank test designed for family-based data. We use genomewide association (GWAS) data funded through the Genetics Analysis Information Network (GAIN) initiative, a public-private partnership between the NIH and the private sector (http://www.fnih.org/GAIN2/home_new.shtml). We consider this an exploratory analysis; the primary analyses of the ADHD diagnostic phenotype and ADHD symptoms have previously been published (Lasky-Su ; Neale and others).

Families were identified through ADHD probands aged 5 to 17 attending outpatient clinics at the data collection sites in Europe and Israel. A total of 958 affected proband-parent trios were initially selected for the GWAS scan. Additional details about the clinical characteristics of this sample have been described in several other manuscripts(Brookes and others 2007; Chen and others 2008; Christiansen and others 2008; Kuntsi and others 2006; Mulligan and others 2008; Zhou and others 2008). All probands had been diagnosed by investigator clinicians has having either ICD-10 or DSM-IV ADHD. Based on the structured interview data, most of the probands met criteria for DSM-IV combined-type ADHD; however a few probands met criteria for DSM-IV inattentive subtype (N=13), hyperactive subtype (N=33) or missed one of the ADHD diagnoses by a single item on the structured interview (N=19). These families were retained in the analysis after reviewing the medical records and structured interview data and confirming the ADHD diagnosis. Family members were of European ancestry from seven countries around Europe including Belgium, Germany, Ireland, the Netherlands, Spain, Switzerland, and the United Kingdom, as well as Israel.

Information was collected on the age at onset of inattentive and hyperactivity-impulsive symptoms using the Parental Account of Childhood Symptom (PACS). The PACS is a reliable, semi-structured interview that measures children's behavior (Taylor and others 1986a; Taylor and others 1986b) which was administered by Master's level investigators at each center to the parents of the affected child. There was centralized training for all who

administered the PACS and the responses to questions were standardized. For both inattention and hyperactivity-impulsivity, the following question was asked of the parent about their child, "How old was X when you first noticed this happening?" The parent then responded by providing the year at which the first symptoms occurred. Therefore, this measure of the age at onset of ADHD (which is how this will be denoted in the manuscript) is not a measure of the official ADHD diagnosis according to specific ADHD criteria, but a measure of when the first ADHD symptoms occurred. Because the validity of this research hinges on the genetic relevance of the age when ADHD symptoms were first observed, we estimated the heritability of this phenotype using a separate sample(2007) of 481 related ADHD individuals where we estimated the heritability of the trait to be 0.19 (p-value = 0.02), suggesting that this trait is genetically relevant. Unfortunately we were unable to do this in the IMAGE data itself because we only have the age at first symptoms recorded for the proband.

Details of the genotyping and data cleaning process were reported elsewhere (Neale and others). Briefly, genotyping was performed by Perlegen Sciences using the Perlegen platform. The Perlegen Array has 600,000 tagging SNPs designed to be in high linkage disequilibrium with untyped SNPs for the three HapMap populations. Genotype data cleaning and quality control procedures were done by The National Center for Biotechnology Information (NCBI) using the GAIN QA/QC Software Package (version 0.7.4) developed by Gonçalo Abecasis and Shyam Gopalakrishnan at the University of Michigan. A copy of the software is available by e-mailing gopalakr@umich.edu or goncalo@umich.edu . Data were removed on the basis of the following quality control metrics: 1) call rate < 95%; 2) gender discrepancy; 3) per-family Mendelian errors >2; 4) sample heterozygosity < 32%; 5) genotype call quality score cut-off < 10; 6) a combination of SNP call rate and minor allele frequency (MAF) (a) $0.01 \leq MAF < 0.05$ and call rate $\geq$ 99%; b) $0.05 \leq MAF < 0.10$ and call rate $\geq$ 97%; and c) $0.10 \geq MAF$ and call rate $\geq$ 95%); 7) Hardy-Weinberg equilibrium p-value of p-value < 0.000001; and 8) duplicate sample discordance.

Family-based association tests (FBATs) use genetic data from family members to evaluate the possible association of a phenotype and a gene allele. In this analysis, we apply a family-based logrank test, FBAT-logrank, to the GWAS data. FBAT-logrank incorporates a commonly used survival analysis approach, the logrank test, into the FBAT test statistic. In this context, the logrank test compares the rates with which individuals of different genotypes are diagnosed with ADHD. FBAT-logrank can be thought of as a standard survival analysis for family data where the age at first symptoms of ADHD indicates the apparent beginning of the disorder. Details on this approach can be found elsewhere(Lange and others 2004). This analysis is available in the PBAT package (http://www.biostat.harvard.edu/~clange/default.htm). In all analyses, we considered additive, dominant, and recessive genetic models. We also examined the association p-values of the SNPs in a set of pre-specified ADHD autosomal candidate genes that was generated by the IMAGE study investigators. These genes are as follows: ADRA1A, ADRA1B, ADRA2A, ADRA2C, ADRB1, ADRB2, ADRBK1, ADRBK2, ARRB1, ARRB2, BDNF, CHRNA4, COMT, CSNK1E, DBH, DDC, DRD1, DRD2, DRD3, DRD4, FADS1, FADS2, HES1, HTR1B, HTR1E, HTR2A, HTR3B, NFIL3, NR4A2, PER1, PER2, PNMT, SLC18A2, SLC6A1, SLC6A2, SLC6A3 (DAT1), SLC6A4, SLC9A9, SNAP25, STX1A, STY1, TPH1, and TPH2. Because a previous association was identified in DRD5 using an independent sample(2007), we attempted to replicate this finding in the IMAGE sample. Unfortunately, there were no SNPs anywhere near DRD5, presumably because DRD5 is part of a known pseudo-gene and therefore very difficult to genotype accurately. This complication prevented us from attempting to replicate our previous finding.

A total of 438,784 markers were available for analytic use after data cleaning procedures. The PBAT/FBAT programs are not compatible with sex-linked markers and therefore we restricted our statistical analysis to 429,981 autosomal markers included among the clean SNPs. A total of 2803 individuals, 1865 founders and 938 non-founders were included after the cleaning process. Eight additional offspring did not have information about the onset of the disease, resulting in 930 individuals used in the genetic analyses. A summary of the sample is listed in table 1. In addition, age at onset had an inverse correlation with the number of hyperactive-impulsive symptoms (r = −0.11, p = 0.0005). Although the direction of the correlation was also inversely related inattentive symptoms, it was not significant (r = −0.038, p = 0.248).

Online supplementary tables 2 and 3 summarize the findings from the analysis. SNPs were excluded from tables 2 and 3 after examination of the survival plots for any of the following reasons: 1) violation of the proportional hazards assumption, as evidenced by crossing of the lines for the different genotypes; 2) there was not good differentiation between the lines for the different genotypes; and 3) the lines for the additive model did not have the heterozygous genotype in between the two homozygous genotypes throughout the survival plot (This occurred less than 10 times among the top results). Online supplementary table 2 displays the association findings with p-values less than 10^{-5}. There are 4 associations with p-values in intronic regions of ADAMTS2 (rs9687070 = 9.34×10^{-6}, 2.98×10^{-6}, rs10039254 = 7.87×10^{-6}, rs3776816 = 4.64×10^{-6}). The lowest 3 association p-values were found in a region of chromosome 6 at SNPs rs806276 and rs9451437 (p-values = 3.38×10^{-7}, 1.02×10^{-7}, and 3.08×10^{-6}) and in chromosome 2 at SNP rs1517484 (p-value = 5.42×10^{-7}). In online supplementary table 2, rs806276 is significant under an additive model with allele 'A' associated with later ADHD onset. Therefore having 0, 1, or 2 genotypes is associated with increasingly later ADHD onset. All of the analysis findings in online tables can be interpreted analogously to the illustration with rs806276.

Online supplementary table 3 displays all of the association p-values in the candidate genes that have p-values less than 0.05. We list all nominally significant results at the candidate genes so that for future replication efforts. Because these SNPs are often in linkage disequilibrium with each other, we display the number of 'distinct regions' within a candidate gene that are represented by the associations in online supplementary table 3. These regions were determined by using the confidence interval criteria by Gabriel et al. (2002). With the exception of SLC9A9 and HTR2A, all of the associations refer to only one region within a given candidate gene. HTR2A and SLC9A9 had associations with in 2 and 6 distinct regions respectively.

We used the age at onset of ADHD in a FBAT-logrank analysis to determine if there are SNPs that can distinguish individuals who are diagnosed with ADHD at different times. A true association would indicate that individuals with a specific genotype(s) have an earlier/ later disease onset than individuals with the other genotype(s). No SNPs in the analysis achieved genomewide significance; however, we found 14 SNPs throughout the genome with association p-values less than 10^{-5}. Among the associations, there are SNPs in two genes, ADAMTS2 and SULF2, neither of which seem like obvious biologic candidates for ADHD. Sulfatase 2 is involved in cell signaling by being a coreceptor for cytokines and numerous heparin-binding growth factors (Dai et al., 2005 [PubMed 16192265])[supplied by OMIM]. ADAMTS2 encodes a part of the ADAMTS protein family and mutations in this gene cause a recessively inherited tissue disorder (http://www.ncbi.nlm.nih.gov/sites/entrez? Db=gene&Cmd=ShowDetailView&TermToSearch=92949&ordinalpos=4&itool=EntrezSys tem2.PEntrez.Gene.Gene_ResultsPanel.Gene_RVDocSum). Neither of these genes appear to be expressed in the brain or have obvious relevance to ADHD.

The most interesting findings in this analysis are the associations found in SLC9A9. Not only does this gene have more associations that any of the other candidate genes, with 6 associations that have a p-value less than 0.05, but these SNPs are not in linkage disequilibrium and represent 6 distinct regions of the gene. It is not surprising that there are several associations in SLC9A9, as there are 180 SNPs in this gene and therefore by chance, one would anticipate many more associations. Therefore, not much can be said for sure about these findings. However, in this dataset, the gene consistently had associations, not only in this study, but also with other quantitative measures of ADHD(Lasky-Su). There were some SNPs in this gene that had nominal associations in this study that were among those with the strongest associations using other quantitative traits(Lasky-Su), including rs130575, rs552655, rs13353224, rs708188, and rs13057533. This finding, along with the biologic evidence for this as a candidate gene(de Silva and others 2003), suggests that SLC9A9 should be followed up in future studies.

There are several limitations of this study. Most notably is the reliance of these results on the validity of the age at onset of ADHD as reported by parents. Since this is a retrospective report, the data are susceptible to recall bias. The parental report of the first occurring symptoms may also be affected other outside sources, such as a teacher mentioning ADHD-like behavior or a parent reading an article describing the characteristics of ADHD. In addition, individuals who have predominantly inattentive symptoms may be more difficult to identify than individuals with more hyperactive-impulsive symptoms, as hyperactive-impulsive symptoms are more easily identifiable in young children. In this study we tried to avoid this complication by using predominantly combined-type ADHD individuals. Despite these drawbacks, based on our initial analysis, there is some evidence that the age at first symptom is heritable. Therefore, even given these disadvantages, there is evidence that using this as a phenotypic trait in genetic analyses is useful.

This analysis used the individual's age at onset of ADHD symptoms as a phenotype in an FBAT-Logrank analysis with the IMAGE GWAS data. Age at onset as a phenotype is important, as this will help indicate whether there are genetic factors that contribute to when a child begins exhibiting ADHD-like behavior, as was suggested in a previously(Lasky-Su and others 2007). Although none of the results reached genome-wide significance, there are some nominally significant results in the candidate genes that may be worth replicating. Following-up the modest findings from this analysis should be done to confirm whether the observed associations represent true genetic effects or are false positives.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-IV. Washington, DC: American Psychiatric Association; 1994. p. 886

Chen WJ, Faraone SV, Tsuang MT. Estimating age at onset distributions: A review of methods and issues. Psychiatric Genetics. 1992; 2:219–238.

de Silva MG, Elliott K, Dahl HH, Fitzpatrick E, Wilcox S, Delatycki M, Williamson R, Efron D, Lynch M, Forrest S. Disruption of a novel member of a sodium/hydrogen exchanger family and DOCK3 is associated with an attention deficit hyperactivity disorder-like phenotype. J Med Genet. 2003; 40(10):733–740. [PubMed: 14569117]

Faraone SV, Glatt S, Su J, Tsuang MT. Three potential susceptibility Loci shown by a genome-wide scan for regions influencing the age at onset of mania. American Journal of Psychiatry. 2004; 161(4):625–630. [PubMed: 15056507]

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, et al. The structure of haplotype blocks in the human genome. Science. 2002; 296(5576):2225–2229. [PubMed: 12029063]

Lange C, Blacker D, Laird NM. Family-based association tests for survival and times-to-onset analysis. Stat Med. 2004; 23(2):179–189. [PubMed: 14716720]

Lasky-Su J. Genome-wide association scan of quantitative traits for Attention Deficit Hyperactivity Disorder identifies novel associations and confirms candidate gene associations.

Lasky-Su J, Biederman J, Laird N, Tsuang M, Doyle AE, Smoller JW, Lange C, Faraone SV. Evidence for an association of the dopamine D5 receptor gene on age at onset of attention deficit hyperactivity disorder. Ann Hum Genet. 2007; 71(Pt 5):648–659. [PubMed: 17501935]

Neale BM, Lasky-Su J, Anney R, Franke B, Zhou K, Chen W, Asherson P, Buitelaar J, Banaschewski T, Ebstein R, et al. Genome-wide Association Scan of Attention Deficit Hyperactivity Disorder.

Taylor E, Everitt B, Thorley G, Schachar R, Rutter M, Wieselberg M. Conduct disorder and hyperactivity: II. A cluster analytic approach to the identification of a behavioural syndrome. British Journal of Psychiatry. 1986a; 149:768–777. [PubMed: 3790876]

Taylor E, Schachar R, Thorley G, Wieselberg M. Conduct disorder and hyperactivity: I. Separation of hyperactivity and antisocial conduct in British child psychiatric patients. Br J Psychiatry. 1986b; 149:760–767. [PubMed: 3790875]

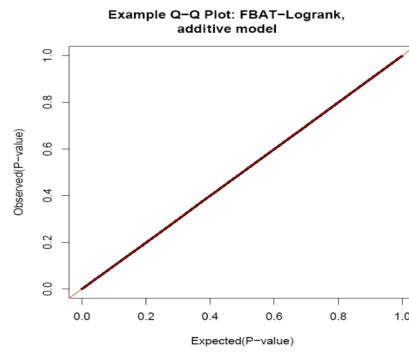**Figure 1.**

Table 1

Descriptive Statistics of the Individuals used in the GWAS analyses.

Number of children	930
Gender distribution of the offspring	
Male (percent)	810 (87.1)
Female (percent)	120 (12.9)
Age among offspring (Standard Deviation)	10.86 (2.82)
Age at onset distribution	
0	7
1	335
2	152
3	146
4	144
5	66
6	68
7	8
>7	4
Average age at onset of ADHD (Standard Deviation)	2.67 (1.73)
Average age at onset of inattention (Standard Deviation)	4.31 (1.96)
Average age at onset of hyperactivity-impulsivity (Standard Deviation)	2.77 (1.85)