Genetic Risk for Attention-Deficit/Hyperactivity Disorder Contributes to Neurodevelopmental Traits in the General Population

Joanna Martin, Marian L. Hamshere, Evangelia Stergiakouli, Michael C. O'Donovan, and Anita Thapar

Background: Attention-deficit/hyperactivity disorder (ADHD) can be viewed as the extreme end of traits in the general population. Epidemiological and twin studies suggest that ADHD frequently co-occurs with and shares genetic susceptibility with autism spectrum disorder (ASD) and ASD-related traits. The aims of this study were to determine whether a composite of common molecular genetic variants, previously found to be associated with clinically diagnosed ADHD, predicts ADHD and ASD-related traits in the general population.

Methods: Polygenic risk scores were calculated in the Avon Longitudinal Study of Parents and Children (ALSPAC) population sample ($N = 8229$) based on a discovery case-control genome-wide association study of childhood ADHD. Regression analyses were used to assess whether polygenic scores predicted ADHD traits and ASD-related measures (pragmatic language abilities and social cognition) in the ALSPAC sample. Polygenic scores were also compared in boys and girls endorsing any (rating $\geq$1) ADHD item ($n = 3623$).

Results: Polygenic risk for ADHD showed a positive association with ADHD traits (hyperactive-impulsive, $p = .0039$; inattentive, $p = .037$). Polygenic risk for ADHD was also negatively associated with pragmatic language abilities ($p = .037$) but not with social cognition ($p = .43$). In children with a rating $\geq$1 for ADHD traits, girls had a higher polygenic score than boys ($p = .003$).

Conclusions: These findings provide molecular genetic evidence that risk alleles for the categorical disorder of ADHD influence hyperactive-impulsive and attentional traits in the general population. The results further suggest that common genetic variation that contributes to ADHD diagnosis may also influence ASD-related traits, which at their extreme are a characteristic feature of ASD.

Key Words: Attention-deficit/hyperactivity disorder, autism spectrum disorder, Avon Longitudinal Study of Parents and Children (ALSPAC), genetics, pragmatic language, social communication

Attention-deficit/hyperactivity disorder (ADHD) is a highly heritable neurodevelopmental disorder characterized by early-onset, developmentally inappropriate inattentive, hyperactive, and impulsive behaviors (1). The disorder occurs more frequently in boys, with a male-to-female ratio of about 3–7:1 (2,3). Similar to other common disorders, the genetic architecture of ADHD is complex, with rare and common variants involved (4). Although clinical diagnoses are defined categorically, ADHD psychopathology can also be viewed dimensionally, with inattentive and hyperactive-impulsive symptoms distributed continuously in the general population (5). Twin and epidemiological studies have shown that heritability estimates for dimensional ADHD are similar across a variety of cutoff points (6,7). This similarity in heritability estimates indicates that genetic factors act throughout the full distribution of ADHD symptoms. However, the postulated relationship between dimensional measures of ADHD in the population and clinical diagnoses has not yet been confirmed at the level of molecular genetics.

It has become clear in more recent years that the boundaries between different neurodevelopmental and psychiatric disorders are not clear-cut, as exemplified by the observed clinical and genetic overlap between ADHD and other disorders. Rates of co-occurrence are especially high for ADHD and autism spectrum disorder (ASD), another highly heritable neurodevelopmental disorder, characterized by social communication and interaction deficits as well as restrictive and repetitive behaviors (8). Studies of children with clinical diagnoses have found that large (>500 kb), rare (<1% frequency) copy number variants in ADHD show significant overlap with copy number variant loci previously implicated in ASD (9,10), although a more recent collaborative cross-phenotype analysis found no clear common genetic overlap in diagnosed ADHD and ASD cases (11). Similar to ADHD, ASD can also be viewed dimensionally (12), and twin studies have found that ADHD and ASD traits share common genetic influences in the general population as well as at the quantitative extreme (13–19). These studies suggest that genetic variants associated with the diagnosis of ADHD might also contribute to population variation in ASD-related trait measures.

Previous research has suggested that children with a clinical diagnosis of ADHD ($n = 452$) differ from control subjects ($n = 5081$) on the basis of a polygenic risk score, an aggregate score of thousands of common alleles of very small effect that together form an index of genetic risk for ADHD (20). In the present study, we tested the hypothesis that en masse common genetic variants that confer risk for a clinical diagnosis of ADHD are associated with ADHD traits in the general population. Given the established clinical and genetic overlap between ADHD and ASD

From the MRC Centre for Neuropsychiatric Genetics and Genomics (JM, MLH, MCO, AT), Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Cardiff; and MRC Integrative Epidemiology Unit (ES), University of Bristol, Bristol, United Kingdom.

Address correspondence to Anita Thapar, F.R.C.Psych., Ph.D., MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Hadyn Ellis Building, Maindy Road, Cardiff CF24 4HQ, UK; E-mail: thapar@cardiff.ac.uk.

Received Aug 22, 2013; revised Feb 12, 2014; accepted Feb 13, 2014.

0006-3223/$36.00
http://dx.doi.org/10.1016/j.biopsych.2014.02.013

(13,14,16), we also analyzed the secondary hypothesis that en masse ADHD common genetic variants are also associated with ASD-related social communication traits in the general population.

Methods and Materials

Target Population Sample

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a large, well-characterized longitudinal data set (21,22). ALSPAC originally recruited pregnant women (N = 14,541) residing in Avon, England, with expected delivery dates of April 1, 1991–December 31, 1992. An additional 713 eligible children whose mothers did not enroll during pregnancy were enrolled after age 7, resulting in a total sample of 14,701 of children alive at age 1 year. Full data (phenotypic and genotypic) were available for up to 5661 children, depending on the outcome variables. Children with >30% missing items on any outcome variable were excluded from analyses of that variable. The study website (http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/) contains details of all available data. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and local research ethics committees.

Phenotypic Measures. Data on ADHD traits were collected when participants were ~7 years, 7 months old, using the parent Development and Well-Being Assessment (DAWBA) (23). For each ADHD item, parents marked boxes to say whether their child showed the behavior; these were coded as follows: 0 for "no," 1 for "a little more than others," and 2 for "a lot more than others." A total ADHD trait score was calculated by summing these responses to give a possible range of 0–36. Scores were also calculated for inattentive and hyperactive-impulsive ADHD traits separately (with a possible range of 0–18 each).

Social communication traits were assessed using the Social and Communication Disorders Checklist (SCDC) (24) and the pragmatic language scales of the Children's Communication Checklist (CCC) (25). A quantitative measure of restricted, repetitive behaviors was not available. Both the CCC and the SCDC have been shown to have good predictive reliability for a clinical diagnosis of ASD in the ALSPAC sample (26). The CCC shows good interrater reliability (.80), internal consistency (.80–.87), and validity for language problems (25), and the SCDC shows good internal consistency (.93), high test-retest reliability (.81), and validity for a diagnosis of ASD (24). The SCDC assesses social cognition and understanding, whereas the CCC pragmatic language scales measure ability to use language in a social context. Previous research has shown that children with ADHD or ASD have lower pragmatic language ability scores than control subjects with typical development, but children with ASD have lower scores than children with ADHD (27).

The SCDC was assessed at the same time as the DAWBA ADHD measures. Parents were asked to judge how much 12 descriptions applied to their child's behavior. The responses were coded as follows: 0 for "not true," 1 for "quite/sometimes true," and 2 for "very/often true." A total SCDC score was calculated by summing these responses (with a possible range of 0–24).

An abridged version of the CCC was used to assess language abilities at ~9 years, 7 months of age. Parents were asked to rate whether statements about their child were "certainly true," "somewhat true," or "not true," which were coded as 0, 1, and 2. The following subscales were summed to generate a pragmatic language abilities score: inappropriate initiation, coherence, stereotyped conversation, conversational context, and conversational rapport. Subscale scores were based on six to eight items each. The pragmatic language total score was obtained for children with data available for each subscale. Because the CCC measures language abilities, lower scores suggest pragmatic language deficits.

Information on DSM-IV ADHD diagnoses is available based on the DAWBA at ~7 years of age. Data on ASD diagnoses are available based on clinical records, using a clinician's diagnosis of ASD (28). Prorated scores were used for measures with <30% missing items.

Genetic Data. After quality control (QC), genome-wide data for 500,527 single nucleotide polymorphisms (SNPs) were available for 8229 of the children, of whom 4213 (51.2%) were boys. Details of QC procedures are provided in Supplement 1.

Discovery Clinical Sample for Generating ADHD Polygenic Risk Scores

The analytic method described by the International Schizophrenia Consortium (29) was used to identify ADHD risk alleles in a discovery genome-wide association study (GWAS) from which polygenic risk scores were derived in the ALSPAC subjects. A published GWAS of British and Irish children with a confirmed DSM-IV research diagnosis of ADHD (n = 727) and population control subjects (n = 5081) was used to define risk alleles. This clinical sample was selected as the primary discovery sample because it is similar to the ALSPAC general population in ethnicity and underwent similar diagnostic assessment procedures. The ascertainment of DNA samples, QC procedures, and GWAS results were described in detail previously (4). This GWAS was based on 502,702 SNPs after strict QC. Following the International Schizophrenia Consortium study, alleles that were more common in cases than controls at SNPs showing evidence for association at the very relaxed threshold $p < .5$ were considered risk alleles.

Generating Polygenic Scores

Full details are available in Supplement 1. In brief, SNPs in approximate linkage equilibrium in the ALSPAC genome-wide data were identified using the PLINK software, available for free download at http://pngu.mgh.harvard.edu/~purcell/plink/ (30). From this set of SNPs, we retained alleles that showed evidence for weak association ($p < .5$) in the discovery ADHD GWAS and used those to calculate a polygenic score for each individual in ALSPAC using PLINK (30). The polygenic scores were standardized using z score transformations.

Data Analysis Strategy

In the ALSPAC sample, children with ADHD or ASD diagnoses were compared with each other and with the remainder of the sample on ADHD, SCDC, and CCC traits, using Student t test. Girls and boys were also compared. Analyses were conducted on the 8229 ALSPAC children with full genetic data available after all QC.

As a result of a strongly negatively skewed distribution of the CCC pragmatic language data, variables were transformed ($\ln x + 1$) and linear regression analyses were performed to test for association with ADHD polygenic score. The ADHD and SCDC traits were highly positively skewed, contained an excess of zero values, and could not be transformed to normality (see Figure 1 for variable distributions). Analyzing such data using standard linear regressions may yield biased estimates of parameters and increased type I and II error rates (31,32). The distribution of data was better explained by a negative binomial than a Poisson distribution of simulated data with the same mean and number (Figure S1 in Supplement 1). These data were analyzed using zero-inflated negative binomial (ZINB) regression models. Gender was included as a covariate in all models.

The ZINB model consists of two submodels that allow for a distribution with an inflated number of individuals with values of zero: 1) logistic regression model of an unobserved dichotomous