STXBP5 Antisense RNA 1 gene and adult ADHD symptoms

Alejandro Arias-Vásquez*, Alexander J. Groffen*, Sabine Spijker*, Klaasjan G. Ouwens*, Marieke Klein*, Dina Vojinovic*, Tessel E. Galesloot, Janita Bralten, Jouke-Jan Hottenga, Peter J. van der Most, V. Mathijs Kattenberg, Rene Pool, Ilja M. Nolte, Brenda W.J.H. Penninx, Iryna O. Fedko, Conor V. Dolan, Michel G. Nivard, Anouk den Braber, Cornelia M. van Duijn, Pieter J. Hoekstra, Jan K. Buitelaar, Bart Kiemeney, Martine Hoogman, Christel M. Middeldorp, Harmen H.M. Draisma, Sit H. Vermeulen, Cristina Sánchez-Mora, J. Antoni Ramos-Quiroga, Marta Ribasés, The EAGLE-ADHD Consortium, Catharina A. Hartman, J.J. Sandra Kooij, Najaf Amin, August B. Smit**, Barbara Franke**, Dorret I. Boomsma**

** These authors contributed equally to this work*

*** These authors shared final responsibility*

This chapter is submitted.

The supplemental information for this paper is available at https://drive.google.com/drive/folders/1POG-CJ_RV52bSRwAtbc4dOJFfM7RAYzs?usp=sharing

ABSTRACT

Attention-deficit/hyperactivity disorder (ADHD) is characterized by age-inappropriate levels of inattention and/or hyperactivity-impulsivity and persists into adulthood in a substantial proportion of cases. ADHD is heritable and is thought to represent the clinical extreme of a continuous distribution of ADHD symptoms in the general population. We aimed to detect ADHD risk conferring genes leveraging the power of population studies of ADHD symptoms in adults. Within the SAGA (Study of ADHD trait Genetics in Adults) consortium, we estimated the SNP-based heritability of self-reported ADHD symptoms and carried out a genome-wide association meta-analysis in nine adult population-based and case-only cohorts of unrelated adults. A total of n = 14,689 individuals were included. We found a significant SNP-based heritability for self-rated ADHD symptom scores of respectively 15% (n = 3,656) and 30% (n = 1,841) in the two cohorts. The top-hit of the genome-wide meta-analysis (SNP rs12661753) was present in the hitherto uncharacterized long non-coding RNA *STXBP5-AS1* gene. This association was also observed in a meta-analysis of childhood ADHD symptom scores in eight population-based pediatric cohorts from the EAGLE ADHD consortium (n = 14,776). Genome-wide meta-analysis of the SAGA and EAGLE data (n = 29,465) increased the strength of the association on the *STXBP5-AS1* gene. In human HEK293 cells, expression of *STXBP5-AS1* enhanced the expression of a reporter-construct of *STXBP5*, a gene known to be involved in SNARE complex formation. In mouse strains featuring different levels of impulsivity, *Stxbp5-AS1* transcript levels in the prefrontal cortex strongly correlated with motor impulsivity as measured in the 5-choice serial reaction time task (r^2 = 0.55). Our results implicate the *STXBP5-AS1* gene in ADHD symptom scores and point to vesicle transport as a biological mechanism involved in ADHD-related impulsivity levels.

INTRODUCTION

Attention-deficit/hyperactivity disorder (ADHD) is a common neurodevelopmental disorder affecting 2–5% of children[1,2] and adults.[3,4] ADHD is characterized by age-inappropriate, sustained symptoms of inattention and/or hyperactivity-impulsivity. In children[5] and adults,[6] ADHD shows substantial heritability. Heritability estimates are largely independent of the phenotypic measurement scale (i.e., categorical or continuous) in children; in adults, estimates are lower when using self-report rating scales.[3] Twin studies[5] suggest that etiological influences on ADHD symptoms are distributed throughout the population, consistent with a liability model.[7] Inattention and hyperactivity-impulsivity symptoms can be reliably assessed in population-based cohorts based on rating scales,[8] creating the possibility to collect large samples for gene-finding studies. The genetic contributions to ADHD in children and adults are complex, with multiple different genetic variants contributing to the disorder,[4] both common and rare.[3] Recently, 16 genome-wide associations have been established in an ADHD Genome-Wide Association Studies meta-analysis (GWASMA) of childhood case-control studies from the Psychiatric Genomics Consortium (PGC) and The Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH) and population-based samples from the EArly Genetics and Lifecourse Epidemiology (EAGLE) consortium.[9,10]

Here, we sought to leverage the power of population studies of ADHD symptoms in adults to detect disease-relevant genes. Within the SAGA (Study of ADHD trait Genetics in Adults) consortium, we estimated the SNP-based heritability of self-reported adult ADHD symptoms and subsequently carried out a GWASMA in nine cohorts of European Caucasian origin (n = 14,689 individuals, age 18 years or older), in whom adult self-reported ADHD symptom scores were available. These samples included six population-based cohorts, two clinical ADHD samples and one clinical cohort ascertained for depressive and anxiety disorders (to enrich the clinical extreme of the ADHD symptom continuum). The locus with the strongest statistical association was followed-up in a replication analysis of quantitative childhood ADHD symptom scores (n = 14,776) from the EAGLE consortium.[9] Genetic correlations were obtained between the PGC and the iPSYCH sample of children[10] and the SAGA sample of adults. Finally, we conducted gene-based tests for genes with SNPs showing a p-value $< 1\times10^{-6}$ in the meta-analysis, making use of the common SNPs from SAGA and rare variant data from the Erasmus Rucphen Family (ERF) study (see **Table 1**), one of the adult cohorts.

Functional follow-up studies downstream of gene-finding in ADHD, e.g. in model systems, to determine the biological relevance of a genetic finding, are scarce.[11] Core features of ADHD, inattention, hyperactivity, and impulsivity are well defined e.g. in mouse

models.[12] Here we carried out functional follow-up studies for the hitherto uncharacterized top-gene of the GWASMA in three mouse inbred strains with large differences in motor impulsivity derived from reaction time tasks, and in a human cell assay.

METHODS

ADHD symptom scores and study populations

ADHD symptom scores were assessed by three instruments (see **Table 1**) in nine cohorts (for a complete description of each sample please follow the references in **Table 1**): the ADHD-index of the Conners Adult ADHD Rating Scale[13] (CAARS ADHD-index; 12 items), the total scores of the DSM-IV ADHD Rating Scale (ADHD-RS),[14] and the Attentional Deficit/Hyperactivity Problems subscale from the Adult Self Report (ASR-ADHD; 13 items).[15] The CAARS (used in NESDA, NTR, ERF) is an extensively tested psychometric instrument with high internal consistency and reliability. Five cohorts (NeuroIMAGE, BIG, IMpACT-NL, VHIR, NBS) collected information using the ADHD-RS,[14] which has high validity in population-based and case samples. For IMpACT-NL and VHIR, only affected individuals were included. One cohort (TRAILS) assessed ADHD problems through the ASR-ADHD (http://www.aseba.org/).[15-17]

Genetic Variant Calling and Quality Control

An overview of genome-wide single nucleotide polymorphism (SNP; for common variants) genotyping, quality control, and imputation is given in **Supplementary Table 1**. Exomes of 1,336 individuals from the ERF population, which is a genetically isolated population in the Netherlands,[18] were sequenced (see **Supplementary Methods**), and ADHD index data were available for 587 of these individuals. Detection of rare variants in the ERF study was done for those genes with SNPs with p-value $< 1\times10^{-5}$ in the GWASMA and variants identified in these exomes were used to estimate the contribution of rare variants in the genes of interest to ADHD behavior (see **Supplementary Methods**).

GCTA

Genome-wide Complex Trait Analysis (GCTA)[19] was used to compute the variance in the ADHD symptom score explained by common SNPs in the two largest cohorts included in the meta-analysis, the NTR and NESDA (n > 1,500 unrelated subjects). A genetic relationship matrix (GRM) for all individuals in the dataset was estimated based on SNPs with high imputation quality (see **Supplementary Methods**). Bivariate GCTA[19] was additionally run on the ADHD-index of the CAARS and ASR-ADHD data also available in the NTR cohort, to assess the genetic correlation (rg) between the two diagnostic instruments.

Genome-wide association and meta-analysis

GWAS was conducted in each cohort by linear regression under an additive model. Age was included as a covariate, but not gender, which was not significantly associated with the ADHD scores in any study. Four principal components were added to account for possible population stratification effects. Information on software packages is provided in **Supplementary Table 1**. In all analyses, the uncertainty of the imputed genotypes was taken into account. Location of SNPs reported is from the build 37 (hg19) 1000G data. Meta-analysis was conducted in METAL (www.sph.umich.edu/csg/abecasis/ metal/ index.html) by the *p*-value-based method, given the intrinsic variability of the quantitative traits used (see **Supplementary Methods**). The meta-analysis was performed in the full sample (nine cohorts) and restricted to the population-based samples (seven cohorts; "restricted sample").

Replication in the EAGLE consortium

Within EAGLE, association of ADHD-related measures was assessed in nine population-based childhood cohorts with genotype data imputed against the 1000 Genomes reference panel.[9] Linear regression of the phenotype on sex, age, genotype dose, and principal components was performed in all cohorts, followed by meta-analysis based on *p*-values in METAL. The TRAILS cohort is part of both consortia and was excluded from the EAGLE consortium for replication analysis, leaving a total of 14,776 children from eight cohorts.

Look-up of significant GWAS loci

Evidence for an effect of the 12 independent ADHD-associated SNPs from the PGC+iPSYCH GWASMA on adult ADHD symptoms was studied through a look-up of results. LD-independent loci with corresponding index-SNPs were obtained from **Table 1** of Demontis *et al*.[10] If the index variant was not present in the SAGA data set, a proxy variant was selected using LDlink (https://analysistools.nci.nih.gov/LDlink/). The Bonferroni-corrected significance level was set at p-value = 0.05/12 = 0.00417.

Linkage disequilibrium score regression (LDSR) analysis

LDSR was used to estimate the genetic correlation between the PGC+iPSYCH sample of children[10] and the SAGA sample of adults. Each dataset underwent additional filtering for markers overlapping with HapMap Project Phase 3 SNPs, INFO score ≥ 0.9 (where available), and MAF ≥ 1%. Indels and strand-ambiguous SNPs were removed. LDSR analysis was performed using the LDSR package (https://github.com/bulik/ldsc[20], see **Supplementary Methods**).

Gene-wide analysis of common and rare variants

Genes containing SNPs with *p*-values < 1×10^{-6} in the meta-analysis of the nine cohorts were selected for gene-wide tests using common and rare variants. The common variant analysis was performed in MAGMA.[21] Flanking regions of 25kb for each gene were included in the analyses. The rare variant analysis was performed with the Sequence Kernel Association Test (SKAT; only in the ERF study) library of the R-software.[22]

Functional analyses

Follow-up functional analyses were performed on the locus containing the best association *p*-value. This locus contains *STXBP5-AS1*, representing a putative long noncoding RNA, predicted to be expressed in several species (**Supplementary Methods**; **Supplementary Table 5** & **Supplementary Fig. 2**). Human *STXBP5-AS1* encodes multiple splice variants, many of which lack a region that overlaps the *STXBP5* gene. To test for regulatory effects of *STXBP5-AS1* on the expression of *STXBP5*, a fluorescent reporter construct was designed to contain the region of antisense overlap (see **Supplementary Methods**).

Mouse models

RNA was derived from prefrontal cortex of adult male mice from the inbred strain C57/Bl6J (n = 7) and recombinant inbred strains BXD29 (n = 8) and BXD68 (n = 7), and gene expression was quantitated (see **Supplementary Methods**). Strains were bred in the facility of the Neuro-Bsik consortium of the VU University (Amsterdam, The Netherlands) and used for behavioral analysis.[12,23]

RESULTS

Quantitative assessment instruments are listed in **Table 1**. The quantitative phenotypes showed a weak, negative correlation with age and no association with gender in any cohort. Phenotypic and genetic correlations between symptom scores assessed with the different instruments were substantial: in a clinical sample of 120 adults with ADHD the phenotypic correlation between the CAARS[13] (ADHD-index) and the ADHD-RS[24] (total score) was high (r = 0.73; *p*-value < 0.01).[24] In 380 parents of children with ADHD, the correlation was of similar magnitude (r = 0.69; *p*-value < 0.001).[25] We estimated the phenotypic correlation between the CAARS ADHD-index and the ASR-ADHD[15,16] in the NTR (n = 15,226; average age 40 years, SD=16.1) to be 0.67 (*p*-value < 0.0001). In younger participants in the age range of the TRAILS cohort (18–22 years, n = 2,687), the correlation was similar (0.68, *p*-value < 0.0001).

Table 1. Descriptive information for all cohorts and for phenotype assessment in the SAGA consortium.

Cohort Name	N (% F)	Age (SD)	Symptom list (N items)	Score range*	Mean Score (SD)*	Ref
NTR	5935 (63%)	43.7 (15.2)	CAARS ADHD-index (12)	0–30	7.9 (3.7)	(Willemsen et al. 2010)
NESDA	1977 (66%)	46.5 (13.0)	CAARS ADHD-index (12)	0–32	8.7 (5.4)	(Boomsma et al. 2008)
ERF	1043 (53%)	45.6 (13.3)	CAARS ADHD-index (12)	0–25	7.8 (4.4)	(Aulchenko et al. 2004)
NeuroIMAGE	470 (51%)	42.3 (5.3)	ADHD-RS (23)	0–43	14.1 (8.9)	(von Rhein et al. 2014)
BIG	448 (63%)	22.3 (3.2)	ADHD-RS (23)	0–40	14.0 (6.4)	(Hoogman et al. 2012)
NBS	2925 (53%)	57.4 (16.3)	ADHD-RS (23)	0–15	1.4 (2.2)	(Galesloot et al. 2017)
IMpACT-NL±	113 (62.8%)	37.7 (11.5)	ADHD-RS (23)	1–18	12.04 (3.3)	(Franke and Reif 2013)
VHIR±	559 (32%)	33.3 (10.6)	ADHD-RS (18)	4–54	31.0 (9.7)	(Bosch R 2019)
TRAILS	1215 (48%)	19.0 (0.6)	ASR ADHD (13)	0–22	5.9 (4.4)	(Ormel et al. 2014)

Conners' Adult ADHD Rating Scale (CAARS ADHD-index), DSM-IV ADHD Rating Scale (ADHDRS), and Attentional Deficit/Hyperactivity Problems subscale from the ASR (ASR ADHD);
*Untransformed values observed per cohort;
± only affected individuals included.

A significant SNP-based heritability was estimated for the CAARS ADHD-index in a subsample of each of the two largest cohorts: 30% (SE = 16.7%, *p*-value = 0.035) in NESDA (n = 1,841 unrelated subjects) and 15% (SE = 7.8%, *p*-value = 0.020) in NTR (n = 3,881 unrelated participants). We also estimated the genetic correlation for the CAARS ADHD-index and the ASR-ADHD using bivariate GCTA. In all individuals from the NTR with genotype and phenotype data (n = 6,036 related and unrelated subjects), the genetic correlation was 0.818 (SE = 0.256). When analyzing the bivariate data in 2,921 unrelated subjects, the point estimate of the genetic correlation was 0.813 (SE = 0.364). The significant SNP-based heritability and the considerable phenotypic and genetic correlations between assessment instruments support the validity of our meta-analysis approach of GWA results obtained across contributing data sets.

For the nine separate GWAS, the genomic control inflation factors (lambda) ranged between 0.996 and 1.026 (mean lambda 1.009, **Supplementary Table 2**). Meta-analysis (**Supplementary Figure 1A**) of the full sample revealed the lowest *p*-value (3.03×10^{-7}) for the intronic SNP rs12661753 in *STXBP5-AS1* (**Supplementary Figure 3**); for the meta-analysis of the restricted sample, *p*-value for this SNP was 1.48×10^{-6} (**Supplementary Figure 1B**). Replication was observed for rs12661753 (*p*-value = 3.07×10^{-2}) for childhood

ADHD symptoms in the EAGLE-ADHD consortium.[9] The subsequent GWASMA between SAGA and EAGLE revealed the best association *p*-value = 2.05×10^{-7} for SNP rs12664716 (n = 29,465; **Supplementary Figure 3F**) located in the *STXBP5-AS1* gene, and in high LD (D' = 1.0, r^2 = 0.98) with rs12661753 (p-value $_{SAGA-EAGLE}$ = 3.55×10^{-7}; **Supplementary Figure 1C**).

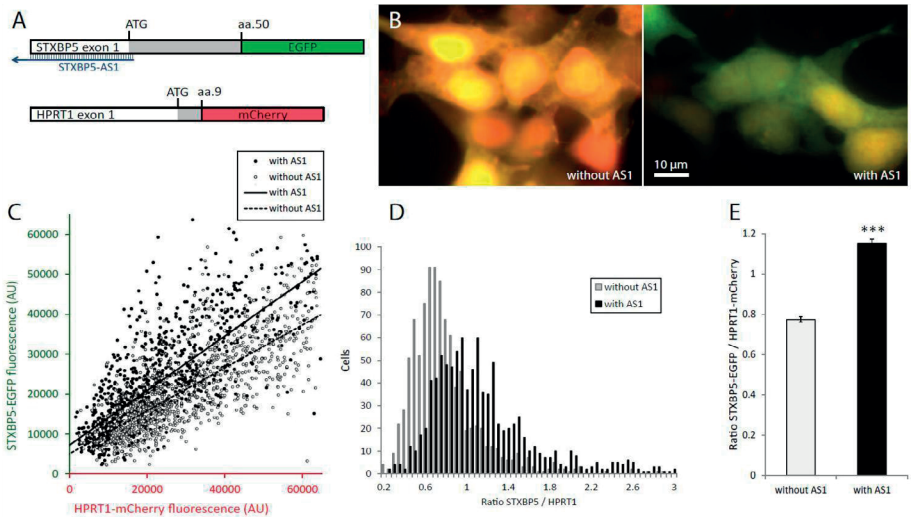


Figure 1. STXBP5-AS1 positively regulates the expression of its cognate mRNA. (A) Design of two reporter constructs. Top: Exon 1 of human *STXBP5*, containing the natural 5'UTR and encoding the first 50 amino acids, was fused in-frame to EGFP. The *Stxbp5-AS1* transcript including the region showing perfect (100%) sequence overlap with the encoded *Stxbp5* transcript is depicted schematically in blue. Bottom: To control for transfection efficiency and differences in cellular metabolic rates, we co-expressed a non-target mRNA comprised of human *HPRT1* exon 1 fused to mCherry. (B) Typical examples of HEK293 cells expressing both constructs with or without STXBP5-AS1. (C) Quantitation of EGFP and mCherry fluorescence in presence or absence of AS1 (947 and 974 cells respectively). (D-E) The ratio of STXBP5-EGFP and HPRT1-mCherry was calculated for each cell. Data are presented as a histogram (D) or as mean±SEM (E). ***, *p*-value = 6×10^{-51}; t_{946} = 4.4412, Student's t-test.

The index variant rs12661753 was not associated with ADHD risk in the recent case-control PGC+iPSYCH GWASMA of ADHD in a sample mainly consisting of children (*p*-value = 0.6316, n = 55,374). A look-up of genome-wide significant ADHD index SNPs from this PGC+iPSYCH GWASMA for association in the SAGA consortium also revealed no significant associations with adult ADHD symptoms (**Supplementary Table 6**).

We estimated the genetic correlation between PGC+iPSYCH and the complete SAGA sample to be 0.541 (SE = 0.447, *p*-value = 2.26×10^{-1}; the VHIR cohort present in both studies). We tested if the two rg values differed significantly from each other, which was not the case (X^2-based test *p*-value > 0.05) (**Supplementary Methods**).

In NTR and NESDA, a subset of participants (n = 6,678) had additional phenotype data on hyperactivity/impulsivity and inattention symptom subscales of the CAARS available. These scales of each 9 items are non-overlapping with the 12 ADHD-index items. For hyperactivity/impulsivity symptoms, the *p*-value for association with rs12661753 was 1.51×10^{-5}, whereas for inattention it was 3.53×10^{-2}, suggesting a stronger effect of the variant on hyperactivity/impulsivity.

As shown in **Table 2** and **Supplementary Table 3**, 50 common variants from 8 independent (clumped) loci showed *p*-value < 1×10^{-6}. Of these, four were also amongst the top-associated loci from the restricted SAGA GWASMA (no patients; **Supplementary Table 4**). The genes closest to these SNPs were selected for gene-wide analysis (**Table 2**). Analysis of common variants in seven genes (plus 25kb flanking regions) in the SAGA GWASMA showed significant association with ADHD symptoms. Two significant findings (*p*-value < 0.003) were for long intergenic non-protein coding RNA genes (*LINC01247, LINC00534*), and nominal significant associations (*p*-value < 0.05 gene-wide) were seen for *STXBP5-AS1*, *CALB1*, *GNG12-AS1*, *STXBP5* (**Supplementary Table 5**). It is important to note that *STXBP5* and *STXBP5-AS1* have no physical separation, thus their 25kb flanking regions overlap. The rare variant analysis also showed nominal association for *STXBP5*. For four genes (*GNG12-AS1, LINC01247, STXBP5-AS1, LINC00534*), rare variants were not observed/detected (**Supplementary Table 5**).

Table 2. Most strongly associated (clumped) SNPs (*p*-value < 1×10^{-6}) coming from the meta-analysis of nine cohorts from the SAGA consortium in physical position order (hg19).

SNP name	Chr	Locus	Pos*	*p*-value	Tested/ Non-Tested Allele	Frequency Tested Allele#	Gene(s) in locus
rs11209188	1	1p31.3	68455306	7.88×10^{-6}	A/G	0.534	*GNG12-AS1*
rs1930272	1	1p31.1	83491910	4.75×10^{-6}	T/C	0.544	*LOC107985037*
rs1564034	2	2p25.2	6510305	2.15×10^{-6}	T/G	0.670	*LINC01247*
rs28734069	4	4q26	120042409	5.77×10^{-6}	T/C	0.016	*LOC102723967; LOC105377395*
rs12661753	6	6q24.3	147409235	3.02×10^{-7}	A/G	0.962	*STXBP5-AS1*
rs13274695	8	8p23.2	3723378	6.00×10^{-6}	A/G	0.077	*CSMD1; LOC105377790*
rs2189255	8	8q21.3	91190297	9.61×10^{-6}	T/C	0.703	*CALB1; LINC00534*
rs73204517	13	13q21.33	69920315	7.19×10^{-6}	C/G	0.126	Downstream *LINC00383*

*bp position based on the GRCh37.p13 build;

#Allele frequency of tested allele based on N = 14,689.

Given that the *STXBP5-AS1* gene, which contains the top-hits, is hitherto uncharacterized, we investigated its function. *STXBP5-AS1* encodes a long noncoding RNA (lncRNA). Although human *STXBP5-AS1* does not have any orthologues listed in the UniGene

database, it is conserved in primates and shows a modest conservation in rodents (**Supplementary Table 7** and alignment in **Supplemental Figure 2**). In the hg19 genome release annotation *STXBP5-AS1* is located next to *STXBP5* in the opposite orientation, with antisense sequence overlap in exon 1 of *STXBP5* (**Figure 1A**). It may be hypothesized that *STXBP5-AS1* affects *STXBP5* expression. For such natural antisense RNAs, both repression and positive effects on the expression of cognate genes have been described.[26,27] We tested this hypothesis by designing a reporter gene fusing exon 1 of *STXBP5* to *EGFP*, and quantifying its expression in human HEK293 cells. Expression of the antisense lncRNA variant *STXBP5-AS1-003* (containing the overlap with *Stxbp5*) caused an increase in the fluorescence ratio between *STXBP5-EGFP* and the control (**Figure 1B-E**).

Given the *in vitro* effects on STXBP5-EGFP protein expression, we tested the relationship between gene expression of mouse *Stxbp5* and/or *Stxbp5-AS1* and measures of behavioral impulsivity. We analyzed gene expression in medial prefrontal cortex of three mouse inbred strains previously described to have large differences in motor impulsivity.[12] Here, we confirmed the strain difference in motor impulsivity between the BXD68, BXD29, and C57BL/6J strains ($F_{2,20}$ = 6.91, *p*-value = 0.005), measured as premature responses in the 5-choice serial reaction time task. In addition, these strains showed differences in errors of omission ($_{F2,20}$ = 5.18, *p*-value = 0.015), but not attention ($F_{2,20}$ = 0.35, *p*-value = 0.771) **(Figure 2A**). In these mice, we detected expression of a mouse *Stxbp5-AS1* transcript in the prefrontal cortex by real-time quantitative PCR, which differed across strains ($F_{2,19}$ = 11.73; *p*-value < 0.001). This transcript showed lowest expression in the most highly impulsive strain, BXD68 (BXD68: 4.58±0.11, C57BL/6J: 5.25±0.14, BXD29: 5.19±0.07, $p\text{-value}_{\text{BXD68 vs C57BL/6J}}$ = 0.003, t_{13} = 3.73; $p\text{-value}_{\text{BXD68 vs BXD29}}$ < 0.001, t_{14} = 4.63) (**Figure 2B**). Expression of *Stxbp5* mRNA was not different between the three strains (BXD68:9.89±0.24; C57BL/6J: 9.83±0.10; BXD29: 9.99±0.23). These results suggest that the role of *STXBP5-AS1* plays in impulsivity is not due to influencing the level of the *STXBP5* transcript. Examining correlations between *Stxbp5-AS1* transcript level and impulsivity/inattention measures, we found a significant correlation with motor impulsivity (r^2 = 0.55; *p*-value = 8.26×10^{-5}, Bonferroni-corrected *p*-value < 0.0083) and a nominally significant association with attention, when measured as errors of omission[28] (r^2 = 0.1765; *p*-value = 5.16×10^{-2}), but not when measured as percentage correct responses (r^2 = 0.0862; *p*-value = 1.85×10^{-1}). Expression of *Stxbp5* did not correlate with these parameters (**Figure 2C**).

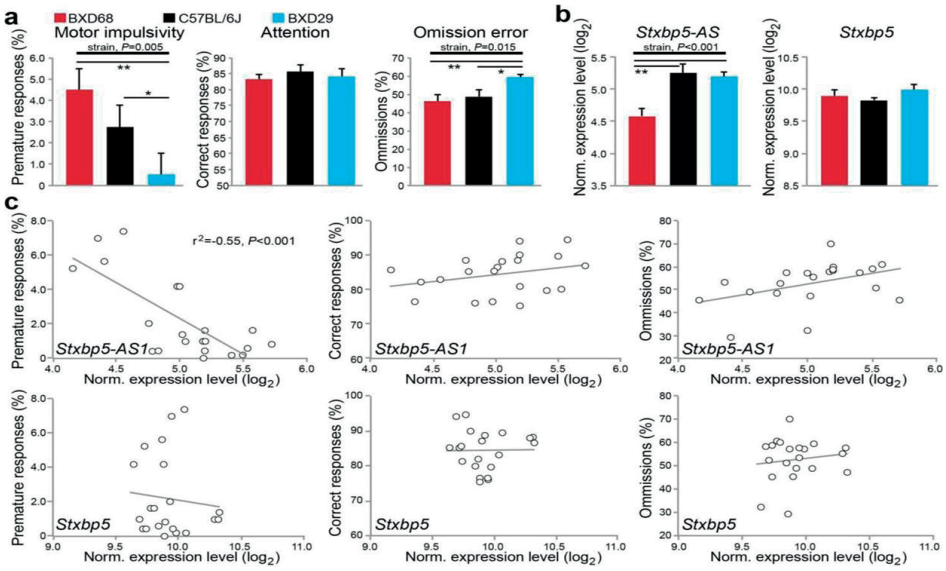


Figure 2. Prefrontal cortex gene expression of putative mouse *Stxbp5-AS1* is correlated with impulsivity. a. Mouse strains BXD68 (red, n = 7), C57BL/6J (black, n = 8) and BXD29 (blue, n = 8) were selected based on a difference in premature responses (motor impulsivity; BXD68 vs. BXD29, t_{14} = 3.71; C57BL/6J vs. BXD29, t_{15} = 2.78) and error of omissions (BXD68 vs. BXD29, t_{14} = 3.54; C57BL/6J vs. BXD29, t_{15} = 2.52), without being different on accuracy (Loos et al. 2014). Shown are data (mean±SEM) of the animals used for gene expression analysis (see b). b. Strain mean±SEM of prefrontal cortex gene expression in BXD68 (red, n=7), C57 (black, n=7), and BXD29 (blue, n=8) for *Stxbp5-AS1* (left) and *Stxbp5* (right). *Stxbp5-AS1* is differentially expressed between strains, with lower expression in BXD68. Yet, Stxbp5 shows no differential expression. No difference in variation was observed. c. Gene expression of *Stxbp5-AS1* (upper panels) in individual mice for which behavioral data was available (BXD68, n = 6; C57BL/6J, n = 7; BXD29, n = 8) correlated well with premature responses (motor impulsivity; left), not with accuracy (attention; middle), and showed a trend towards correlation with errors of omissions (right; *p*-value = 0.0516). *Stxbp5* expression (lower panels) did not correlate with any of these parameters. Trend lines are given in gray. **p*-value < 0.05; ***p*-value < 0.01; ****p*-value < 0.001.

DISCUSSION

We report a genetic variant associated with three different but correlated adult ADHD symptom lists in a meta-analysis of nine European adult population-based and case-only cohorts (n=14,689 individuals). The *STXBP5-AS1* gene (best SNP *p*-value = 3.02×10^{-7}) was the most strongly associated locus in a meta-analysis. This association was confirmed in the EAGLE meta-analysis ($p\text{-value}_{EAGLE} = 2.89\times10^{-2}$), and the top-hit from the full SAGA-EAGLE GWASMA was also located in the *STXBP5-AS1* gene and in almost perfect LD with the original finding (SNP rs12664716, $p\text{-value}_{SAGA\text{-}EAGLE} = 2.05\times10^{-7}$; n = 29,465).

For the adult ADHD-index, an earlier large twin family study estimated total heritability at 30%, and common SNPs thus contain substantial information concerning its genetic

variance. SNP-based heritability analyses, which were ran prior to GWASMA, provided estimates of 15-30% explained variance of adult ADHD symptom scores in the general population. Such estimates are comparable with the estimates obtained for ADHD and four additional categorically defined psychiatric disorders,[29] providing rationale for a gene-finding enterprise for adult ADHD symptoms in the general population.

The function of the *STXBP5-AS1*-encoded lncRNA is currently unknown. *STXBP5-AS1* has been proposed as a prognostic biomarker for survival of cancer patients,[30] but no information is available for its role in ADHD, related traits, or other psychiatric diseases. It overlaps in anti-sense with *STXBP5* encoding a protein involved in synaptic function by regulating neurotransmitter release through stimulating SNARE complex formation.[31,32] This complex plays a major role in intracellular vesicular trafficking in eukaryotic cells and is involved in the exocytotic release of neurotransmitters during synaptic transmission.[33] Genes related to the SNARE complex and its regulators have been investigated in ADHD,[34] and current results suggest that this complex may exert distinct roles throughout development, with age-specific effects of its genetic variants on ADHD behavior.[35] Specifically, deletions and mutations of *STXBP5* occur in autism[36] and epilepsy.[37] *STXBP5* has a presynaptic role that negatively regulates neurotransmitter release by forming syntaxin-SNAP25-tomosyn complex.[38] However, the postsynaptic role of STXBP5 has not been well elucidated.

Post-hoc analysis suggested that *STXBP5-AS1* affects hyperactivity-impulsivity more strongly than inattention. The stronger link with impulsivity was corroborated in behavioral studies in mice. Our experiments in HEK293 cells showed that the lncRNA does not cause antisense inhibition of *Stxbp5*. The increased fluorescence of a reporter protein containing mouse *Stxbp5* exon 1, together with unchanged *Stxbp5* mRNA levels in mouse strains expressing different *Stxbp5-AS1* levels, suggest that the lncRNA might enhance Stxbp5 protein translation or stability. Alternatively, *Stxbp5-AS1* might contribute to impulsivity by a Stxbp5-unrelated mechanism. In line with this idea, *Stxbp5-AS1*, expression (but not that of *Stxbp5*) correlated negatively with motor impulsivity in mice.

Our study should be viewed in the light of some strengths and limitations. A pro was the sample size that could be achieved for quantitative data available through a population-based approach. Moreover, the functional analyses provided a very strong candidate associated with adult and childhood ADHD symptoms. A limitation of our study was the combination of three different phenotyping instruments, but given the strong phenotypic and genetic correlations between the instruments, this might not have reduced power substantially.

The genetic correlation of PCG+iPSYCH with SAGA should be interpreted carefully because the standard error is high. The fact that the PCG+iPSYCH/SAGA rg (0.54), did not differ from the published rg estimate between the PCG+iPSYCH GWASMA and a GWAS of the 23andMe sample (0.65, SE = 0.114)[10] is encouraging but not unexpected given the low power to detect a difference. The estimated genetic correlation between the 23andMe and PGC+iPSYCH analyses was significant but lower than the genetic correlation of the EAGLE and PCG+iPSYCH childhood cohorts (rg = 0.943, SE = 0.204, *p*-value = 3.65×10^{-6}).[10] The ADHD diagnosis (yes/no) in 23andMe is based on the self-reported answer to a single question about presence of a lifetime diagnosis of ADHD[10] and we do not know if the 23andMe participants were diagnosed in childhood or as adults. With a further increase in GWAS sample size update rg results could suggest that there are different genetic correlation patterns between the association results estimated from the GWAS of adult (population-based) ADHD behavior and the GWAS from children, at this point the lack of power makes these analyses inconclusive.

Our study shows that self-reported adult ADHD symptoms measured in the general population have a genetic component and that performing population-based GWASMA of adult ADHD symptoms provides novel insights into the genetic underpinnings of hyperactivity/impulsivity symptoms that are a hallmark of ADHD. Our findings implicate synaptic function regulation through *STXBP5-AS1* and potentially *STXBP5* in ADHD symptom etiology.

REFERENCES

1. Polanczyk, G. & Rohde, L.A. Epidemiology of attention-deficit/hyperactivity disorder across the lifespan. *Curr Opin Psychiatry* **20**, 386-92 (2007).
2. Association, A.P. *Diagnostic and statistical manual of mental disorders*, (Washington, DC, 1994).
3. Franke, B. *et al.* The genetics of attention deficit/hyperactivity disorder in adults, a review. *Mol Psychiatry* **17**, 960-87 (2012).
4. Faraone, S.V. *et al.* Attention-deficit/hyperactivity disorder. *Nat Rev Dis Primers* **1**, 15020 (2015).
5. Faraone, S.V. *et al.* Molecular genetics of attention-deficit/hyperactivity disorder. *Biol Psychiatry* **57**, 1313-23 (2005).
6. Saviouk, V. *et al.* ADHD in Dutch adults: heritability and linkage study. *Am J Med Genet B Neuropsychiatr Genet* **156B**, 352-62 (2011).
7. Caspi, A. *et al.* A replicated molecular genetic basis for subtyping antisocial behavior in children with attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry* **65**, 203-10 (2008).
8. Larsson, H. *et al.* Genetic and environmental influences on adult attention deficit hyperactivity disorder symptoms: a large Swedish population-based study of twins. *Psychol Med* **43**, 197-207 (2013).
9. Middeldorp, C.M. *et al.* A Genome-Wide Association Meta-Analysis of Attention-Deficit/Hyperactivity Disorder Symptoms in Population-Based Pediatric Cohorts. *J Am Acad Child Adolesc Psychiatry* **55**, 896-905 e6 (2016).
10. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for ADHD. *bioRxiv* (2017).
11. Klein, M. *et al.* Brain imaging genetics in ADHD and beyond - Mapping pathways from gene to disorder at different levels of complexity. *Neuroscience and Biobehavioral Reviews* **80**, 115-155 (2017).
12. Loos, M. *et al.* Neuregulin-3 in the Mouse Medial Prefrontal Cortex Regulates Impulsive Action. *Biological Psychiatry* **76**, 648-655 (2014).
13. Conners, C.K. Clinical use of rating scales in diagnosis and treatment of attention-deficit/hyperactivity disorder. *Pediatric Clinics of North America* **46**, 857 (1999).
14. Kooij, J.J.S. *et al.* Reliability, Validity, and Utility of Instruments for Self-Report and Informant Report Concerning Symptoms of ADHD in Adult Patients. *Journal of Attention Disorders* **11**, 445-458 (2008).
15. Kessler, R.C. *et al.* Validity of the World Health Organization Adult ADHD Self-Report Scale (ASRS) Screener in a representative sample of health plan members. *International Journal of Methods in Psychiatric Research* **16**, 52-65 (2007).
16. Kessler, R.C. *et al.* The World Health Organization adult ADHD self-report scale (ASRS): a short screening scale for use in the general population. *Psychological Medicine* **35**, 245-256 (2005).
17. Achenbach, T.M. & Rescorla, L.A. *Manual for the ASEBA Adult Forms & Profiles*, (Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families, 2003).
18. Aulchenko, Y.S. *et al.* Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet* **12**, 527-34 (2004).
19. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
20. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).

21. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *Plos Computational Biology* **11** (2015).
22. Wu, M.C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* **89**, 82-93 (2011).
23. Spijker, S. *et al.* Morphine exposure and abstinence define specific stages of gene expression in the rat nucleus accumbens. *FASEB J* **18**, 848-50 (2004).
24. Sandra Kooij, J.J. *et al.* Reliability, validity, and utility of instruments for self-report and informant report concerning symptoms of ADHD in adult patients. *J Atten Disord* **11**, 445-58 (2008).
25. Thissen, A.J., Rommelse, N.N., Altink, M.E., Oosterlaan, J. & Buitelaar, J.K. Parent-of-origin effects in ADHD: distinct influences of paternal and maternal ADHD on neuropsychological functioning in offspring. *J Atten Disord* **18**, 521-31 (2014).
26. Kimura, T. *et al.* Stabilization of human interferon-alpha1 mRNA by its antisense RNA. *Cell Mol Life Sci* **70**, 1451-67 (2013).
27. Matsui, K. *et al.* Natural antisense transcript stabilizes inducible nitric oxide synthase messenger RNA in rat hepatocytes. *Hepatology* **47**, 686-97 (2008).
28. Guillem, K. *et al.* Nicotinic acetylcholine receptor beta2 subunits in the medial prefrontal cortex control attention. *Science* **333**, 888-91 (2011).
29. Lee, S.H. *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics* **45**, 984 (2013).
30. Guo, W. *et al.* Transcriptome sequencing uncovers a three-long noncoding RNA signature in predicting breast cancer survival. *Sci Rep* **6**, 27931 (2016).
31. Sakisaka, T. *et al.* Dual inhibition of SNARE complex formation by tomosyn ensures controlled neurotransmitter release. *J Cell Biol* **183**, 323-37 (2008).
32. Yizhar, O. *et al.* Tomosyn inhibits priming of large dense-core vesicles in a calcium-dependent manner. *Proc Natl Acad Sci U S A* **101**, 2578-83 (2004).
33. Antonucci, F. *et al.* SNAP-25, a Known Presynaptic Protein with Emerging Postsynaptic Functions. *Front Synaptic Neurosci* **8**, 7 (2016).
34. Bonvicini, C., Faraone, S.V. & Scassellati, C. Attention-deficit hyperactivity disorder in adults: A systematic review and meta-analysis of genetic, pharmacogenetic and biochemical studies. *Mol Psychiatry* **21**, 872-84 (2016).
35. Cupertino, R.B. *et al.* SNARE complex in developmental psychiatry: neurotransmitter exocytosis and beyond. *Journal of Neural Transmission* **123**, 867-883 (2016).
36. Davis, L.K. *et al.* Novel copy number variants in children with autism and additional developmental anomalies. *Journal of Neurodevelopmental Disorders* **1**, 292-301 (2009).
37. Dhillon, S., Hellings, J.A. & Butler, M.G. Genetics and mitochondrial abnormalities in autism spectrum disorders: a review. *Curr Genomics* **12**, 322-32 (2011).
38. Sakisaka, T. *et al.* Regulation of SNAREs by tomosyn and ROCK: implication in extension and retraction of neurites. *J Cell Biol* **166**, 17-25 (2004).

SUPPLEMETARY METHODS AND TABLES

Supplementary Methods

Supplementary Table 1. Quality control, filtering, pre-imputation, and imputation algorithms used.

Supplementary Table 2. Number of SNPs tested after imputation and the lambda values per sample.

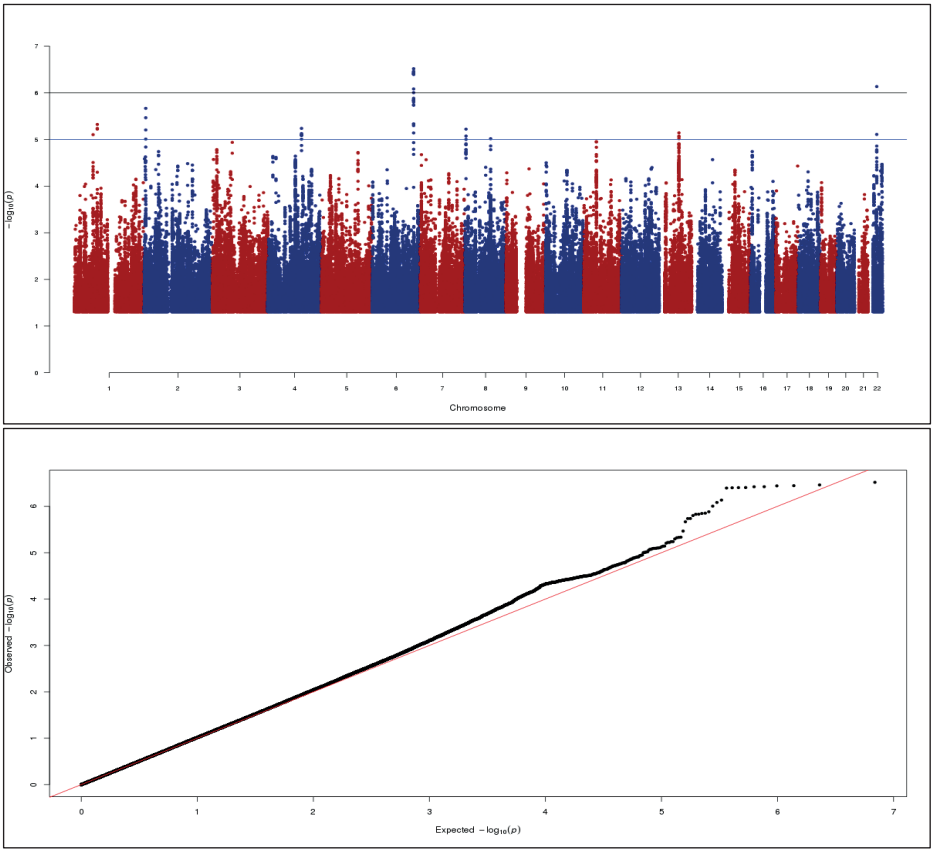
Supplementary Table 3. Most strongly associated SNPs (p-value < 1×10^{-5}) coming from the meta-analysis of nine cohorts from the SAGA consortium in physical position order (hg19).

Supplementary Table 4. Association results for most strongly associated (clumped) SNPs (p-value < 1×10^{-6}) coming from the meta-analysis of seven cohorts from the SAGA consortium without including patients in physical position order (hg19).

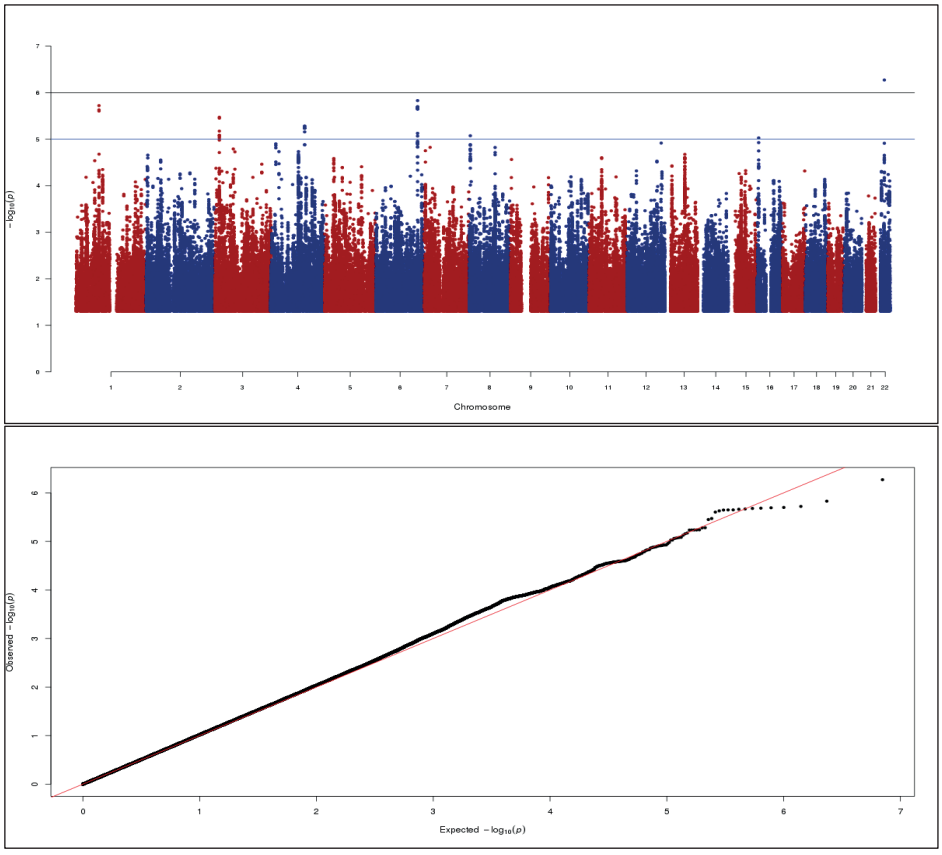
Supplementary Table 5. Gene-wide association p-values for common and rare variants present in the top loci from the SAGA meta-analysis of ADHD Symptom Total Score.

Supplementary Table 6. Best associated loci from the PGC+iPSYCH ADHD GWASMA and their association in the SAGA GWASMA.

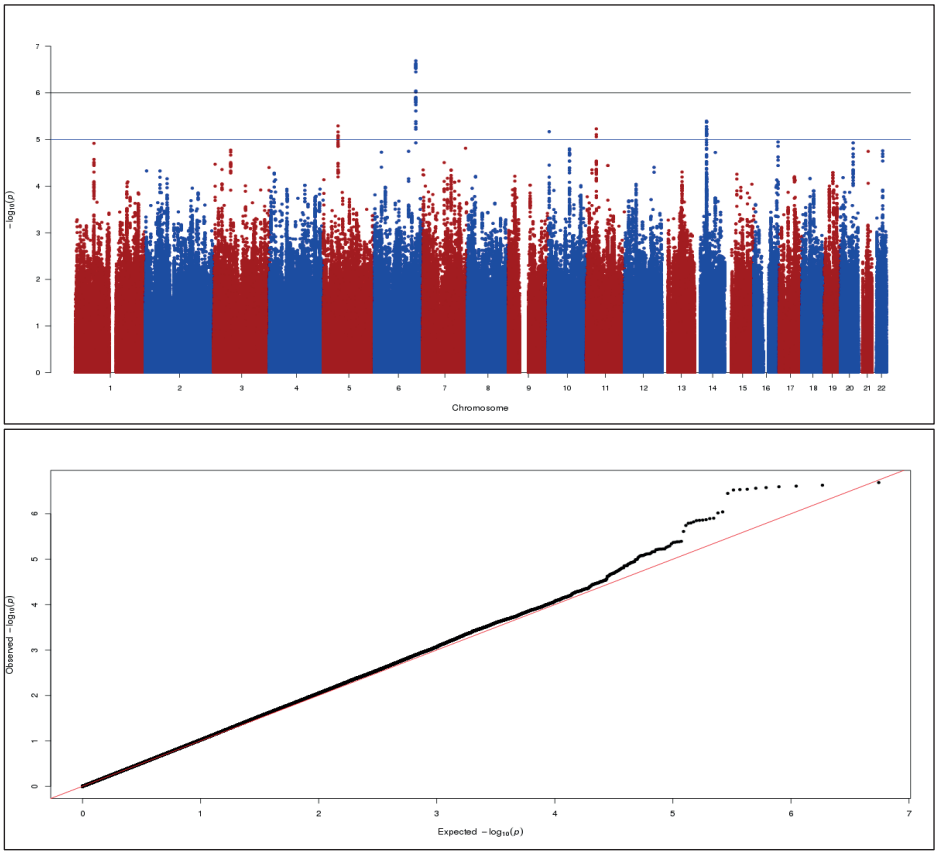
Supplementary Table 7. Conservation and mapping of genomic sequences in primates with similarity to exons encoding human *STXBP5-AS1-003* (Ensembl accession# ENST00000427394).



Supplementary Figure 1A. Manhattan & QQ plot of the ADHD Symptom Total Score meta-analysis from the complete SAGA consortium.



Supplementary Figure 1B. Manhattan & QQ plot of the ADHD Symptom Total Score meta-analysis from the SAGA consortium without patient cohorts.



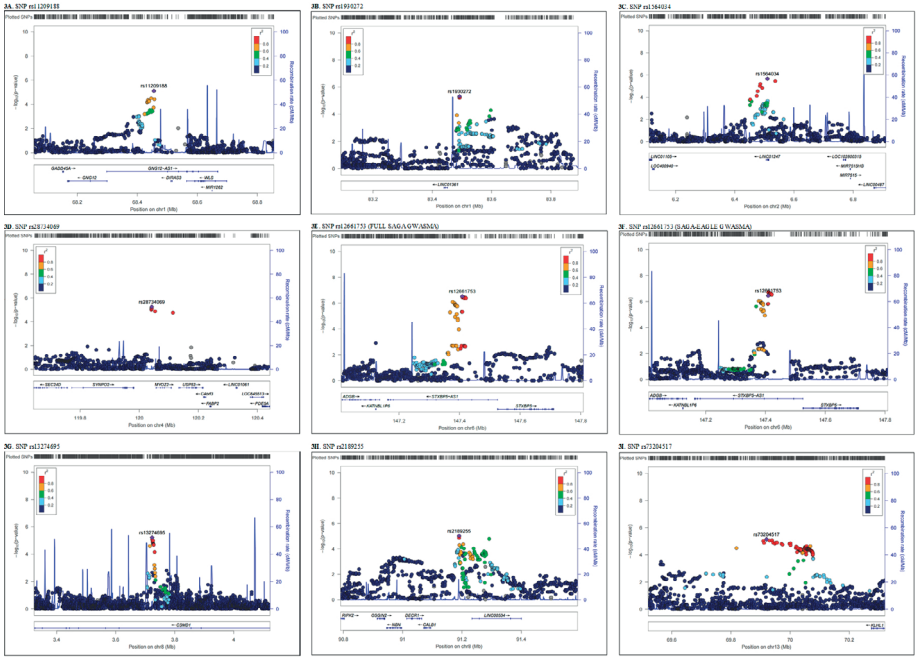
Supplementary Figure 1C. Manhattan & QQ plot of the ADHD Symptom Total Score meta-analysis from the SAGA & EAGLE consortia.

STXBP5-AS1: Exon 6

```
man          agATGCCACTGTTGCTGGAGAAATGTCACTGCTAGAGAACAGAAGACTATTATT
chimp        agATGCCACTGTTGCTGGAGAAATGTCACTGCTAGAGAATAGAAGACTATTATT
gorilla      agATGCCACTGTTGCTGGAGAAATGTCACTGCTAGAGAATAGAAGACTATTATT
orangutan    agATGCCACTGTTGCTGCAGAAATGTCACTGCTAGAGAATAGAAGACTATTATT
rhesus       agATGCCACTGTTGCTGGAGAAATGTCACTGCTAGAGAATAGAAGACTATTATT
gibbon       agATGCCACTGTTGCTGGAGAAATGTCACTGCTAGAGAATAGAAGACTATTATT
baboon       agATGCCACTGTTGCTGGAGAAATGTCACTGCTAGAGAATAGAAGACTATTATT
gr.monkey    agATGCCACTGTTGCTGGAGAAATGTCACTGCTAGAGAATAGAAGACTATTATT
mus          agATGTTGCTGTTGCTGGAGAAATGTCACTGGGAAAGATAA-AGGACCATTATT
rat          agATGTTACTGTTGCTGGAGAAATGTCACTATGAAAGATAA-AGGACCATTATT

man          CCAAAGGGTATTTGAGACTGACTGAATCAGGTCTGGAACATTATTGAAATGgt
chimp        CCAAAGGGTATTTGAGACTGACTGAATCAGGTCTGGAACATTATTGAAATGgt
gorilla      CCAAAGGGTATTTGAGACTGACTGAATCAGGTTTGGAACATTATTGAAATGgt
orangutan    CCAAAAGGTATTTGAGACTGACTGAATCAGGTCTGGAACATTATTGAAATGgt
rhesus       CCAAAGGGTATTTGAGACTGACTGAATCAGGTCTGGAATATTATTGAAATGgt
gibbon       CCAAAGGGCATTTGAGACTGACTGAATCAGGTCTGGAACATTATTGAAATGgt
baboon       CCAAAGGGTATTTGAGACTGACTGAATCAGGTCTGGAATATTATTGAAATGgt
gr.monkey    CCAAAGGGTATTTGAGACTGACTGAATCAGGTCTGGAATATTATTGAAATGgt
mus          CTAAAGAGTGTTTCAGGCTGACTGAATCAGGTCTGAAACGTTATTGAAATGgg
rat          CTAAAGAGTGTTTCAGGCTGACTGAATCAGGTCTGAAACCTTATTGAAATGgt
```

Supplementary Figure 2. ClustalW2 alignment of primate and rodent genomic sequences sharing similarity with human *STXBP5-AS1*. Splice acceptor and (possible) donor sites are indicated in lowercase font. See **Supplementary Table 5** for exon start/end positions and sequence accession numbers. The locations of mouse real-time PCR primers, as used in **Fig. 3**, are indicated in bold.



Supplementary Figure 3. Locus Zoom plots for SNPs from the SAGA Meta-analysis associated with Total ADHD Symptom Score