

Common Genetic Risk Factors in ASD and ADHD Co-occurring Families

Anbo Zhou¹, Xiaolong Cao^{1,5}, Vaidhyanathan Mahaganapathy¹, Marco Azaro¹, Christine Gwin¹, Sherri Wilson¹, Steven Buyske², Christopher W. Bartlett³, Judy F. Flax¹, Linda M. Brzustowicz^{1,4}, Jinchuan Xing^{1,4,*}

¹ Department of Genetics, Rutgers, The State University of New Jersey, Piscataway, NJ, USA

² Department of Statistics, Rutgers, The State University of New Jersey, Piscataway, NJ, USA

³ The Steve Cindy Rasmussen Institute for Genomic Medicine, Battelle Center for Computational Biology, Abigail Wexner Research Institute at Nationwide Children's Hospital; Department of Pediatrics, College of Medicine, The Ohio State University, Columbus, Ohio, USA

⁴ The Human Genetics Institute of New Jersey, Rutgers, The State University of New Jersey, Piscataway, NJ, USA

⁵ Current address: Division of Laboratory Medicine, Zhujiang Hospital, Southern Medical University, Guangzhou, China

* To whom correspondence should be addressed.

Jinchuan Xing

Department of Genetics, Human Genetics Institute of New Jersey

Rutgers, the State University of New Jersey,

145 Bevier Road

Piscataway, New Jersey 08854.

Email: jinchuan.xing@rutgers.edu;

Running title: Risk genes in ASD and ADHD co-occurring families

Abstract

Autism spectrum disorder (ASD) and attention-deficit/hyperactivity disorder (ADHD) are two major neurodevelopmental disorders that frequently co-occur. However, the genetic mechanism of the co-occurrence remains unclear. The New Jersey Language and Autism Genetics Study (NJLAGS) collected more than 100 families with at least one member affected by ASD. NJLAGS families show a high prevalence of ADHD and provide a good opportunity to study shared genetic risk factors for ASD and ADHD. The linkage study of the NJLAGS families revealed regions on chromosomes 12 and 17 that are significantly associated with ADHD. Using whole genome sequencing data on 272 samples from 73 NJLAGS families, we identified potential risk genes for ASD and ADHD. Within the linkage regions, we identified 36 genes that are associated with ADHD using a pedigree-based gene prioritization approach. *KDM6B* (Lysine Demethylase 6B) is the highest-ranking gene, which is a known risk gene for neurodevelopmental disorders, including ASD and ADHD. At the whole genome level, we identified 207 candidate genes from the analysis of both small variants and structure variants, including both known and novel genes. Using enrichment and protein-protein interaction network analyses, we identified gene ontology terms and pathways enriched for ASD and ADHD candidate genes, such as cilia function and cation channel activity. Candidate genes and pathways identified in our study provide a better understanding of the genetic etiology of ASD and ADHD and will lead to new diagnostic or therapeutic interventions for ASD and ADHD in the future.

Introduction

Autism spectrum disorder (ASD) and attention-deficit/hyperactivity disorder (ADHD) are two neurodevelopmental disorders (NDDs) of high prevalence and severity. ASD is characterized by deficits in social interaction and social communication, and by restricted, repetitive, and stereotyped patterns of behavior, interests, and activities (1). ADHD is characterized by inattention and hyperactive/ impulsive symptoms (1). The two disorders have shown a high comorbid frequency where 20-50% of children with ADHD meet the criteria for ASD and 30-80% of ASD children meet the criteria for ADHD (2). Both conditions can have a severe negative impact on patients' quality of life, and only worsen when co-occurring.

On the behavioral level, and especially regarding diagnosis and clinical care, it is often difficult to distinguish some of the behavioral characteristics that define ASD to those that define ADHD. Individual with ASD often suffer from symptoms consistent with ADHD. Lack of focus, impulsivity, and emotional outbursts may be present in both disorders yet the approach to intervention, both behaviorally and pharmacologically, can be substantially different. Clarifying the genetic relationship between these two disorders could inform future treatment decisions. If ASD and ADHD represent pleiotropic manifestations of the same disorder, a unified treatment approach might be effective to target the symptoms of both. If they stem from different etiologies, separate treatments for each might prove more effective. Therefore, a deep understanding of the genetic etiology of familial co-occurrence of ASD and ADHD is important to enhance treatment planning.

Extensive research has been conducted on both disorders and demonstrates overlapping genetic factors between the two conditions (1-7). For example, ADHD candidate causal genes *DRD3* and *MAOA* are cautiously positively associated with ASD (2). In genome-wide association studies (GWAS), 16 single nucleotide variants (SNVs) related to ADHD were found to be possibly involved in ASD and 25 SNVs related to ASD are possibly involved in ADHD (2). In copy number variation (CNV) studies, CNVs segregated with ADHD were also found to be enriched for ASD candidate genes. Correspondingly, ASD patients' family members also carry ADHD diagnosed CNVs (8, 9). While the studies have suggested shared genetic risk factors as one possible reason for the co-occurrence of the two disorders, they either used microarray or whole exome sequencing data. In addition, those studies are based on the case-control study design and do not leverage the power of family studies. Therefore, the evidence of shared genetic risk factors between ASD and ADHD from these studies is mostly weak and inconclusive.

To further explore the genetic etiology of ASD and ADHD co-occurrence, we aimed to use whole-genome sequencing (10) analysis to identify SNVs and structural variants (SVs) responsible for the ASD or ADHD phenotypes segregated in the New Jersey Language and Autism Genetics Study (NJLAGS) project (11). NJLAGS is a project studying the genetic etiology of ASD and related NDDs and has conducted WGS on 272 samples from 73 families, each with at least one ASD patient. All family members have been characterized by extensive behavioral assessments of language and social functioning, restrictive and repetitive behaviors, and other co-occurring behaviors (11). Despite the original study's focus on ASD, these families exhibit

elevated rates of ADHD, with more than 40% of the ASD patients also diagnosed with ADHD. The co-occurrence of ASD and ADHD in the NJLAGS families provides an opportunity to identify genetic factors underlying the co-occurring ASD and ADHD.

Patients and Methods

Sample collection and phenotype assessment

The sample in this study includes 79 families from a previous NJLAGS study (11) (Wave 1) plus an additional 32 families collected after the 2014 publication (Wave 2). The original aim of the NJLAGS study was to identify genetic variation related to DSM-IV autistic disorder, language impairment, as well as related disorders. Families were required to have at least one person with autistic disorder and one additional person with a language-learning impairment (11). During Wave 1, autism probands were required to have a strict diagnosis of autistic disorder based on the Autism Diagnostic Interview (ADI-R), Autism Diagnostic Observation Schedule (ADOS), and the Diagnostics and Statistical Manual-IV (DSM-IV). A second proband had to meet the criteria for Specific Language Impairment (SLI), a disorder where language development is delayed or deviant and cannot be explained by any other neurodevelopmental diagnosis. For Wave 2, NJLAGS sought to understand if the original strict autistic disorder diagnostic criteria were necessary, and therefore, the autism proband requirements were intentionally relaxed to become more in line with the less restrictive, newer DSM-5 criteria for Autism Spectrum Disorder (12). All subjects gave informed consent or assent conforming to the guidelines for treatment of human subjects from the Institutional Review Board at Rutgers, The State University of New Jersey (IRB number: 13-112Mc).

For ASD probands, ADHD characteristics were identified during the intake interview by either a study child psychiatrist or a developmental pediatrician. In each case, the diagnosis was corroborated in detailed NJLAGS history questionnaires that

asked the parents if their child had ever received a diagnosis of attention deficit disorder (ADD) and/or ADHD: 1) Medical History Questionnaire, 2) Family History Questionnaire, and 3) Language Correlates Questionnaire. For all other family members, parents completed the same detailed Family History Questionnaire and Language Correlates Questionnaire regarding whether they or their children have received a diagnosis of several different neurodevelopmental or neuropsychiatric disorders, including ADD/ADHD. In all questionnaires ADD and ADHD were treated as one category and we used ADHD as a collective term for the affected individual.

Genotyping and sequencing

DNA extraction was performed by RUCDR (Piscataway, NJ), either from blood DNA (WB) or Lymphoblastoid cell lines (LCL). Wave 1 Affymetrix Axiom array genotyping experiments have been described previously (11). For Wave 2, single nucleotide polymorphism (SNP) data was generated with the Illumina Infinium PsychArray-24 v1 array (Illumina, San Diego, CA), which includes 593,260 SNPs. SNP with a population minor allele frequency (MAF) >1% were selected. Quality control on SNP genotypes was conducted by array batch and by array type, as described previously (11). The quality control criteria include individual/SNP genotype completion, relationship checking, Mendelian errors checking, and ancestry inference. The linkage analysis only included samples that clustered with the CEU (Utah residents with Northern and Western European ancestry from the CEPH collection) samples from the HapMap reference data, as determined by EIGENSTRAT using the recommend parameters in the documentation (13).

Initially, a subset of 10,899 SNPs in common across the Wave 1 and 2 arrays was selected for the linkage analysis, which minimize marker-to-marker linkage disequilibrium while retaining a high MAF (>30%) to provide suitable genomic coverage of recombination events in the pedigrees. Overlap in some genomic regions was too low to retain acceptable information content as measured by MERLIN (14). In those regions, array-specific SNPs were included. This procedure did not pose an issue with missing data within families since every family was genotyped using only a single array type.

The whole-genome sequencing was performed in four batches by three vendors (Table S1). All samples were sequenced using Illumina paired-end format with a spec of 30x coverage of the human genome. Some samples were excluded in downstream analyses because of quality issues. A few were dropped because the subjects withdrew from the study. For samples that were sequenced in more than one batch, the best quality run was used for the analysis. The raw sequencing reads, variants, and genotypes for all samples are available in the National Institute of Mental Health Data Archive (NDA) under experiments C1932 and C2933.

Linkage Analysis

Linkage analyses were conducted with KELVIN (v2.3.3). KELVIN implements the posterior probability of linkage (PPL) metric to estimate the probability that a genetic location is linked with a tested trait (15). Primary linkage analysis of the phenotypes was conducted on each wave separately, and the linkage evidence was sequentially updated across the waves using Bayes' rule to provide a single metric for linkage evidence. A secondary linkage analysis was conducted using all families jointly in a single pooled analysis of each trait. By comparing the sequentially updated result to the

pooled result, the role of heterogeneity in the dataset can be inferred qualitatively.

Because stratifying on an irrelevant trait will on average produce the same result as a pooled analysis (16), if the sequentially updated PPLs were appreciably higher than pooled results, heterogeneity demarcated by wave is likely to present in the data.

Based on previous simulations of the null distribution in the NJLAGS sample when correcting for three phenotypes (11), a PPL of 0.32 or greater is consistent with a genome-wide error rate of $p < 0.001$, a PPL of 0.26 corresponds to $p < 0.01$, and a PPL of 0.11 corresponds to $p < 0.05$. These threshold values are similar to previous studies of the false positive rate of the PPL after correction for testing multiple phenotypes (17, 18).

Small variant (SNV/indel) and SV calling

Alignment of paired-end FASTQ files was performed using the BWA-MEM algorithm (v0.7.12) (19) to the Human Genome Reference Consortium Build 37 (hg19) using default parameters. The output was converted to BAM format using SAMtools view (v0.1.19) (20). BAM files from read alignment were then processed using the GATK v3.5.0 variant calling pipeline following the best practice recommendation for alignment processing and variant calling (21, 22). Starting from sorted and indexed individual BAM files, a series of GATK alignment processing procedures were conducted, including PCR duplicate removal and base quality score recalibration. SNVs and small insertion/deletions (indels) were called per individual using HaplotypeCaller before being jointly called by GenotypeGVCF. All samples from different sequencing batches were jointly called along with the 1000 Genomes project European ancestry samples (CEU, GBR, FIN) as controls to reduce the batch-effect for downstream analysis (23).

After variant call, we employed variant quality score recalibration using VariantRecalibrator and ApplyRecalibration as outlined in the GATK protocol.

SV calling has been historically difficult for tools using a single algorithm (24). Therefore, an ensemble algorithm, MetaSV (25), along with its components (Breakseq2 (26), breakdancer (27), CNVnator (28), and Pindel (29)) were used for SV discovery. Local realignment by SPAdes (30) and AGE (31) were carried out to further improve breakpoint resolution. MetaSV then combined all produced evidence into a single call set. The outputs of different chromosomes from the same individual were then merged back to one file using VCFtools (32). Basic statistics for the SV calls were calculated by SURVIVOR (33) for quality control. SVs were then filtered to select SVs that had consensus calls from at least two SV callers in MetaSV.

Variant annotation and selection

SNVs and indels were annotated by VAT in the VAAST package (v2.0.2) (34) and condensed into CDR files to represent one family per file by VST in the VAAST package. The variants were filtered to include those that have an MAF < 5% in the ExAC dataset excluding psychiatric cohorts (ExAC non-psych v0.3.1 <https://gnomad.broadinstitute.org/>). For control samples, 635 GTEx WGS samples were obtained from the GTEx project (dbGaP accession number phs000424.v7.p2) (35) and condensed into a single group. SV calls were annotated by AnnotSV (v3.0.9) (36).

Gene prioritization

For SNVs and indels, the gene prioritization tool pVAAST (v0.02) was used to prioritize candidate genes (37). A pVAAST score was calculated for each gene from its variants' linkage pattern, association evidence, MAF, and functional prediction. pVAAST

analyses were performed for the two ADHD linkage regions and the whole genome, under both dominant and recessive modes of inheritance. The linkage region analyses were performed at 10^6 permutations per gene and the whole-genome analyses were performed at 10^5 permutations per gene to determine the significance. For running pVAAST, the pedigrees were processed and trimmed to include individuals from one pair of ancestral parents per pedigree for the dominant mode, or a two-generation subset of the pedigree for the recessive mode. The individuals were selected to maximize the number of sequenced and affected samples in each family.

Candidate genes expressed in at least one of the brain expression databases were selected for downstream analysis. For linkage regions, candidate genes that had a Bonferroni-corrected p-value < 0.05 were considered significant and selected. For whole-genome analyses, candidate genes that have a p-value $\leq 10^{-5}$ and a LOD score ≥ 1 in the pVAAST analysis were selected for downstream analysis.

For SVs, AnnotSV annotation provides a ranking score for potential pathogenicity (36). SVs that are predicted to be likely pathogenic (class 4) or pathogenic (class 5) were selected. In each individual, the severity score of an SV in class 4 and 5 is calculated as: severity score = (class - 3) * number of SV copies, where the number of copies is set at 1 for heterozygous SVs, and 2 for homozygous SVs. A custom script was used to convert results from an SV-based report into a gene-based report. For each gene, the annotated severity scores of all SVs overlapping a given gene were aggregated for affected individuals and unaffected individuals separately. The genes were ranked based on the aggregated SV severity scores in affected individuals. Genes were selected for downstream analysis using the following criteria: 1. containing SV(s)

in more than one affected individual; 2. expressed in at least one of the brain expression databases; 3. SV(s) overlapping the exonic region; and 4. the aggregated score in affected individuals is higher than that in unaffected individuals.

Gene annotation, pathway and enrichment analyses

For gene function predictions, the pLI (probability of being loss-of-function intolerant) score was extracted from gnomAD (v3.0) for each gene (38). Brain developmental gene expression data were obtained from the GTEx project (v8) (35), the BrainSpan Atlas of the Developing Human Brain project (RNA-Seq GENCODE v10 summarized to genes, <http://www.brainspan.org/static/download.html>) (39), and the Human Developmental Biology Resource (<https://www.ebi.ac.uk/gxa/experiments/E-MTAB-4840/Downloads>) (40), as previously described (41). The genes' association with specific cells or diseases are predicted by the HumanBase (<https://hb.flatironinstitute.org>) algorithm that learns from large genomics data collections (42). The gene knock-out mouse behavior was downloaded from the International Mouse Phenotyping Consortium (IMPC) (<ftp://ftp.ebi.ac.uk/pub/databases/impc/latest/>) (43). Candidate genes for several NDDs were collected from previous studies and the gene sets and genes are summarized in Table S2.

The protein-protein interaction (PPI) networks were generated based on gene interactions from three databases, including STRING (v11.5) (44), GIANT (v2.0) (42) and ConsensusPathDB (v35) (45). The cutoff for the “combined score” for STRING was set to 700. For GIANT, interactions with score > 0.6 in tissues from brain and neuron system were used, including brain, astrocyte_of_the_cerebellum, astrocyte_of_the_cerebral_cortex, basal_ganglion, brainstem, central_nervous_system,

central_nervous_system_pericyte, cerebellar_cortex, cerebellum, cerebral_cortex, cerebral_hemisphere, corpus_striatum, forebrain, frontal_cortex, glial_cell, global, globus_pallidus, gray_matter, gray_matter_of_telencephalon, gyrus, hindbrain, hippocampal_formation, hypothalamus, lobe_of_cerebral_hemisphere, neocortex, nervous_system, neural_cell, neuron, nucleus_of_brain, pallium, parietal_lobe, peripheral_nervous_system, pituitary_gland, pons, spinal_cord, substantia_nigra, telencephalic_nucleus and telencephalon. The Python NetworkX package (<https://networkx.org>) was used to visualize the PPI network and the positions of genes in the network were saved for final plotting with R (v4.1.0) packages ggplot2 and scatterpie. Details of interaction data download and processing procedures were described previously (41, 46).

Gene set enrichment analyses were performed using over-representation analysis provided by ConsensusPathDB (45). All pathway, Gene Ontology (GO), and protein complex-based enrichment analyses were enabled, with the minimum two genes from the input and p-value cutoff set to 1. Enrichment p-values for all genes were determined by ConsensusPathDB using a hypergeometric test and enrichment p-value for each gene list was calculated with Fisher's exact test as described previously (41). An enriched term (*i.e.*, GO, pathway, or protein complex) was selected for further analysis if: 1) the total number of genes belong to the term is ≤ 500 ; 2) the term includes more than three genes from all input genes (ADHD, ASD/ADHD, and NDD genes); 3) enrichment p-value < 0.01 for ADHD and/or ASD/ADHD gene lists.

Results

Overall study design

For samples in the NJLAGS cohort, we conducted the SNP genotyping microarray analysis and the WGS analysis in parallel (Figure 1). Genotypes from the genotyping microarray were used to perform linkage analysis and the WGS data were then used to prioritize genes based on SNVs/indels and SVs in both linkage regions and the whole genome. To account for the genetic heterogeneity of ASD and ADHD, candidate gene sets were subjected to enrichment and pathway analyses.

Among the 73 families involved in WGS, 47 (64.3%) have individuals who are also affected by ADHD. For the 493 total individuals within the 73 families, 125 are affected by either ASD or ADHD (Table 1). Among the 98 individuals affected by ASD, 41 (41.8%) are also affected by ADHD (Table 1). Males are more likely to be affected by both ASD and ADHD compared to females (male/female ratio: 2.73, T-test $p = 0.03$).

Linkage analysis highlights linkage peaks for ADHD

A total of 524 individuals from 111 families were genotyped for the linkage analysis. The fully constructed pedigrees contain 707 individuals, after adding missing individuals (e.g., grandparents) to link sibships of cousins. Families have an average of 5.7 individuals (including ungenotyped individuals) and 4.2 genotyped individuals.

We performed multipoint linkage analysis of the data for three phenotypes: “ADHD”, where individuals diagnosed with ADHD are considered affected; “ADHD or ASD” (referred to as “ASD/ADHD” in the following text), where both ADHD and ASD affected individuals are considered affected; and “ADHD and ASD”, where only individuals diagnosed with both ADHD and ASD are considered affected. The results

are summarized in Figure 2 with peaks summarized in Table S3. The ADHD phenotype was linked to 17p13.1-2, meeting the conventional standard for declaring linkage (PPL=0.38, $p < 0.001$) as determined by a simulation study to estimate the empirical null distribution (Figure 2). Because the pooled PPL was larger than the sequentially updated PPL, we infer that the locus is largely homogenous and these results do not offer evidence that data from either of the two waves are inconsistent with linkage. As such, the ADHD locus did not depend on the strictness of the ASD criteria used in family ascertainment. The locus on 12q12-15 met the criteria for suggestive linkage to ADHD (PPL=0.27, $p < 0.01$, Figure 2). Similar to the locus on 17p, the pooled PPL for 12q was larger than the sequentially updated PPL, indicating that the linkage with ADHD did not differ across the two waves with different ASD recruitment criteria. Several additional loci met the criteria of nominal linkage. The ADHD phenotype is nominally linked to 3p13 (pooled PPL=0.11, sequential PPL=0.23) with evidence for heterogeneity across ASD criteria and 19q12-13.1 (pooled PPL=0.20, sequential PPL=0.14) with no evidence for heterogeneity. The phenotype ADHD/ASD had two nominal linkage peaks, both suggesting heterogeneity, on 19p13.3 (pooled PPL=0.11, sequential PPL=0.18) and 20q13.13-13.33 (pooled PPL=0.02, sequential PPL=0.19). The “ADHD and ASD” phenotype did not show evidence for linkage, likely due to the small sample size of the phenotype (Figure 2).

Whole genome sequencing and variant discovery

We selected 272 individuals from the recruited families to undergo WGS to ~30x coverage. The sequencing reads that passed quality filters were mapped to the human

genome reference hg19 and genomic variants are called jointly using GATK for small variants (SNVs and indels) and MetaSV for SVs.

The number of small variants and SVs were consistent across samples. For small variants, around 4 million SNVs were identified for most of the samples, and the number of indels ranged from 700,000 to 1.2 million (Table S1). For SVs, an average of 743,280 SVs (including unknown type) were discovered for each individual, ranging from 325,443 to 1,291,816 across samples. We selected SVs that were identified by at least two methods for the downstream analysis. An average 1,802 SVs passed the filter in each individual (see Methods for detail), ranging from 1,332 to 2,620 (Table S1). Most SVs are within the 50-1,000 bps range and the largest category is deletion (Figure S1).

Candidate genes associated with ADHD and ASD/ADHD in linkage regions

To identify the risk genes for ADHD, we first examined SNVs and indels within the chromosome 12 and 17 significant ADHD linkage regions using the pVAAST workflow. pVAAST prioritizes genes by integrating segregation information, population allele frequency, and functional impact of variants in a gene (37). After Bonferroni correction of the p-values (1,299 genes for chromosome 12 and 563 for chromosome 17 linkage regions) and selecting genes express in brain, we identified 34 genes that met the significance criteria (Bonferroni-corrected $p < 0.05$) for the dominant mode and 3 genes for the recessive mode (Table 2). Proline glutamate and leucine rich protein 1 (*PELP1*) is present in both lists. Among the 36 unique genes, 18 are located in the chromosome 12 linkage region and 18 in the chromosome 17 linkage region, respectively. Four genes (*KDM6B*, *COL2A1*, *PTPRB*, *PER1*) were reported in one or more NDD disease gene databases (Table 2). In addition to SNVs and indels, we also identified SVs using

the WGS data. Within the linkage regions, an average of 83 SVs were identified per individual, with a minimum of 32 and a maximum of 157 across the cohort. None of the genes passed our SV gene filtering criteria (see whole-genome SV analysis below for filtering detail).

The highest-ranking gene is Lysine Demethylase 6B (*KDM6B*) on chromosome 17. *KDM6B* has two missense variants segregating in two families. One mutation (17-7751888-C-G, p.Thr761Ser) is predicted to be possibly damaging by PolyPhen2. The other (17-7749972-G-T, p.Val209Leu) is predicted to be benign by PolyPhen2. Three additional missense variants are also present in patients but do not follow strict segregation patterns (Table S4). *KDM6B* is expressed in the brain and it is intolerant to mutations (pLI=1). The *KDM6B* protein demethylates trimethylated lysine-27 on histone H3. Pathogenic alterations in histone lysine methylation and demethylation genes have been associated with multiple NDDs (47), including ADHD (48, 49). *KDM6B* is also annotated in NDD databases such as SFARI and iPSYCH, and it is identified as one of the 102 risk genes in a recent large ASD study (50).

The highest-ranking gene in the chromosome 12 linkage region is lysine methyltransferase 2D (*KMT2D*), which is also involved in lysine methylation, similar to *KDM6B*. Two missense variants segregate in two families (12-49428694-T-C, p.Asp3419Gly; 12-49436073-C-T, p.Asp1970Asn), both are predicted to be damaging by both SIFT and PolyPhen2. Nine additional missense variants are present in patients but do not follow strict segregation patterns (Table S4). *KMT2D* is expressed in the brain and intolerant to mutations (pLI=1).

Another candidate gene from the chromosome 17 linkage region, Neuralized E3 Ubiquitin Protein Ligase 4 (*NEURL4*), has a missense variant (17-7224433-C-T, p.Gly1120Arg) segregating in one family and four additional missense variants in other affected individuals. *NEURL4* is expressed in the brain and intolerant to mutations (pLI~1). *NEURL4* protein is a scaffold protein, which maintains normal centriolar homeostasis and prevents formation of ectopic microtubule organizing centers (51). Although no publication directly links *NEURL4* with NDDs, evidence has shown that microtubule related genes are related to neuronal migration and dendritic functioning (52). Knockout of *NEURL4* in mice causes neurological phenotypes such as decreased prepulse inhibition (43).

In the chromosome 12 linkage region, we also identified the circadian rhythm controller gene Timeless Circadian Regulator (*TIMELESS*) with a missense mutation (12-56827209-C-A, p.Ala129Ser). The mutation is segregated in two families and is predicted to be deleterious by SIFT, suggesting its potential role in ADHD etiology. Sleep disorders are common in both ASD and ADHD patients (53). A previous study found mutations in circadian-relevant genes, including *TIMELESS*, are more frequent in ASD affected individuals than controls (54).

Novel and known risk genes associated with ADHD and ASD/ADHD in the whole genome analysis

Because ASD and ADHD are highly heterogeneous and multiple variants/genes could contribute to their genetic etiology, next we performed pVAASST analysis on SNVs/indels for the whole genome. After selecting genes with evidence of inheritance (LOD score \geq 1) and express in the brain (see Methods for detail), we identified 16, 25, 70,

and 29 candidate genes for ADHD dominant model, ADHD recessive model, ASD/ADHD dominant model, and ASD/ADHD recessive model, respectively (Table S5).

From SVs, we identified 209 and 227 genes that contain exonic SVs in at least one affected individual for ADHD and ASD/ADHD, respectively (Table S6). To further prioritize genes affected by SVs, we calculated the combined severity scores of SVs within each gene and determined their brain expression levels. We then selected genes whose combined severity scores were larger in affected individuals than in unaffected ones and were expressed in the brain. A total of 43 and 44 genes passed our SV gene filtering criteria for ADHD and ASD/ADHD, respectively (Table S6).

Next, we created two high-confidence candidate gene sets for ADHD and ASD/ADHD. For ADHD, we combined significant linkage-region candidate genes, high-confidence candidate genes from the whole-genome pVAASST analyses, and the SV analysis. For ASD/ADHD, we combined candidate genes from whole-genome pVAASST and SV analyses. A total of 207 unique genes were included in these two confidence gene sets, including 113 genes for ADHD and 138 for ASD/ADHD (Table 3). We annotated these genes using public resources, including gene association databases, literature search databases, tolerance to loss of function mutation scores, mouse knock-out experiments, and expression databases (Table S7, see Methods for detail). A total of 31 genes (15%) were described by previous studies or existing NDD gene databases (Table S2), including 22 genes from the SFARI database. Three genes (*ANKRD11*, *KDM6B*, *COL2A1*) overlap risk genes in the iPSYCH study of ASD and ADHD (55).

In addition to genes overlapping known NDD genes, many other candidate genes showed functional relevance to neurodevelopmental processes or have been reported

to be associated with NDDs. For example, we identified Protein disulfide-isomerase A2 (*PDIA2*) in the whole-genome pVAASST analysis. With the ADHD dominant model, *PDIA2* had a missense mutation (16-334899-G-A, p.Ala188Thr) segregating in two families and was predicted as damaging by SIFT. It had additional 27 variants segregating in 13 families. *PDIA2* encodes a member of disulfide isomerase family of endoplasmic reticulum (ER) proteins that catalyze protein folding and thiol-disulfide interchange reactions and is highly expressed in the cerebellar cortex. HumanBase predicted a strong association of *PDIA2* to ASD (rank 1) and a study investigating the role of ER stress in autism identified *PDIA2* upregulation in autistic subjects (56).

In whole-genome SV analysis, we discovered a novel candidate gene Calcium Voltage-Gated Channel Auxiliary Subunit Alpha2delta 4 (*CACNA2D4*). *CACNA2D4* is one of the genes that encode the alpha-2/delta subunit of the voltage-dependent calcium channel, among three other genes (*CACNA2D1*, *CACNA2D2*, *CACNA2D3*) (57). While a dozen genes encoding the alpha1 and beta subunits of the calcium channel are previously identified in the SFARI database (Table S2), *CACNA2D4* was absent. In our study, *CACNA2D4* was found to contain SVs in affected individuals in five families in the ASD/ADHD analysis (one SV overlaps intron 16 to intron 26 and other SVs within intron 26). The SV overlapping intron 16 to intron 26 is a 36 kb deletion (chr12:1948911-1985036) that is rare in general populations (<0.16% in European populations in gnomAD, https://gnomad.broadinstitute.org/variant/DEL_12_124891?dataset=gnomad_sv_r2_1). Two previous ASD studies identified *CACNA2D4* deletions in ASD patients and support our finding (58, 59).

Candidate genes participate in pathways related to neurological disorders

To explore relationships among the candidate genes and to identify potential shared mechanisms between ASD, ADHD, and other NDDs, we performed enrichment analysis for the 207 candidate genes with 1,629 known NDD genes (Table S2) and determined the significance of enrichment for each gene list (Table S8). ADHD and ASD/ADHD showed strong enrichment in GO terms and pathways related to cilium and microtubule, suggesting that many genes in this group are involved in brain development (Table S8). For example, pathway “Ciliopathy” (WP4803) is significantly enriched in both ADHD and ASD/ADHD gene sets, overlapping five and seven genes in the two sets, respectively. “cilium or flagellum-dependent cell motility” (GO:0001539) is one of the top enriched GO term in ADHD gene list, and overlap four genes in the candidate genes (*BBS4*, *CCDC65*, *TEKT1*, *TEKT5*).

Other processes that are known to relate to NDDs are also present. For example, “cation channel activity” (GO:0005261) is enriched in both ADHD and ASD/ADHD gene sets, overlapping seven and eight genes in the two sets, respectively. Besides shared enriched terms and pathways, each gene set also contains phenotype-specific enrichments. For example, striatum development (GO:0021756) and cerebral cortex development (GO:0021987) are significantly enriched in the ASD/ADHD gene set, overlapping two and four candidate genes, respectively. On the other hand, “sensory perception of light stimulus” is only significantly enriched in ADHD gene set (GO:0050953), overlapping seven genes in the set.

Protein-protein interaction network of candidate genes

To further examine the relationship among the candidate genes, we constructed PPI networks using our candidate genes and known NDD genes (Figure 3, Table S9). Of the 113 ADHD genes, 23 were connected with known NDD genes into a single PPI network with interactions supported by at least 2 databases (Figure 3A, 45 genes in total), and 50 of the remaining genes were connected with known NDD genes with interactions supported by at least one database (Figure 3B, 150 genes). Similarly, for the 138 ASD/ADHD genes, 44 were connected after adding NDD genes with interactions supported by ≥ 2 databases (Figure 3C, 71 genes in total), and 31 of the remaining genes were connected after adding NDD genes with interactions in ≥ 1 database (Figure 3D, 83 genes). Several enriched terms, such as histone methyltransferase complex and cation channel activity, showed extensive connection with known genes for ASD, ADHD, and other NDDs, supporting the shared genetic etiology of ADHD and ASD.

Discussion

ASD and ADHD co-occur at a high frequency and studies suggested shared genetic risk factors between the two disorders (1-7). Given the worsened negative impact on patients' quality of life when the two conditions co-occur, it is crucial to understand the underlying genetic risk factors for each disorder and the co-occurrence. In the NJLAGS cohort, ADHD occurs in 41.8% of ASD patients, consistent with previous studies (reviewed in (60)). The family design of the NJLAGS study provides a unique opportunity to identify inherent and *de novo* genetic factors underlying both ASD and ADHD.

In this study, we identified two linkage peaks (12q12-15, 17p13.1-2) for the ADHD phenotype from 111 families. Using WGS data from 272 selected samples, we discovered 36 significant genes within the linkage regions and explored their gene functions. Expanding the analysis to the whole genome revealed more risk genes for both ADHD and ASD/ADHD phenotypes. Notably, both top candidate genes in the two linkage regions (*KDM6B*, *KMT2D*) are involved in the histone lysine methylation process. Histone modification is a central process regulating gene expression during neurodevelopment and mutations in *KDM6B* have been linked to multiple NDDs (47), including ADHD (48, 49). At the whole genome level, 31 of the high-confidence candidate genes were identified in previous studies or NDD gene databases. In addition, we observed multiple NDD-related pathways overrepresented in the high-confidence ADHD and ASD/ADHD gene sets (Table S8). Taken together, the results demonstrate the power of combining linkage analysis, rare variant analysis, and SV analyses to uncover risk genes.

By studying ADHD affected patients in families with ASD probands, we identified genetic risk factors underlying ASD and ADHD co-occurrence. Three candidate genes (*ANKRD11*, *KDM6B*, *COL2A1*) were reported in a previous study of ASD and ADHD comorbidity in Danish cohorts (55). In the Danish cohorts, all three genes contain multiple protein-truncating variants or missense variants in ASD or ADHD patients, and *ANKRD11* (Ankyrin Repeat Domain Containing 11) is one of the top candidate genes with a significant excess of constrained rare protein-truncating variants (crPTVs) (55). In addition, several genes (*KDM6B*, *PTPRB*, *PER1*) that are significant within the ADHD linkage region are described in previous ASD studies (61-63). Pathways we discovered that are important in ADHD etiology are also vital in ASD etiology. By providing an extensive candidate gene set, we are not only reaffirming the linkage between ASD and ADHD, but also suggesting novel potential risk pathways for further investigation.

Several important GO terms and pathways recurred in both gene sets in our enrichment analysis, namely cilium/microtubule, ion channel, histone methylation, and chromatin remodeling. These pathways revealed interesting aspects of mechanisms in ASD and ADHD etiology as an interconnected network of hundreds of genes and pathways (64) and underlined the complexity of genetic casualties of NDDs. A recent article argued that several NDDs share common genetic risk factors and there might not be "Autism-specific" genes (65). Our results support this view. We also identified some important but underappreciated processes such as the circadian rhythm (GO:0007623), which has phenotypic effects in cognition, mood, and reward-related behaviors. Research documented that the prevalence of insomnia ranges from 50% to 80% in ASD patients, compared to 9–50% in age-matched typically developing children (66).

However, the underlying mechanism of association and genes involved remain to be determined. The genes related to circadian rhythm we identified in this study (*MYBBP1A*, *PER1*, *TIMELESS*) could shed light on the genetic etiology of sleep disorders in ASD and ADHD patients.

Our study has a few limitations. The limited sequencing capacity led to the sequencing of only a selected number of ADHD affected individuals, as the NJLAGS dataset was collected primarily to identify ASD and language-impaired probands. However, recently we sent out follow-up questionnaires to all NJLAGS families requesting an update on ADHD diagnoses. We specifically asked who had received a diagnosis from a medical professional (*i.e.*, psychiatrist, developmental pediatrician, neurologist, psychologist). Of the 34 families who responded to our questionnaire, 15 were families who had some sequencing data and whose data are included in this manuscript. From these 15 families, 28 individuals had received a formal diagnosis of ADD/ADHD by a medical professional. Importantly, only 13 of these family members had been initially identified as having a diagnosis of ADD/ADHD. In some cases these were younger family members who were too young at the time of data collection to be identified as affected while in other cases, we did not have this diagnostic information available. These updated findings will be included in future ADHD in genetic analyses. Moving forward, we anticipate extensive sequencing efforts to continue for the NJLAGS project resulting in the sequencing of all NJLAGS families. As more ADHD affected individuals are identified and sequenced, this more complete pedigree information and stronger statistical power will allow us to further discover potential common risk factors underlying the ASD and ADHD etiology.

Another limitation of this study is we focused mainly on variants affecting protein coding sequences. Non-coding regulatory variants are known to play an important role in NDDs, although predicting their functional impact is still difficult (67). With the rapid development of functional prediction algorithms for non-coding variants, we will identify additional candidate genes affected by regulatory variants using the cohort's WGS data in the near future.

In conclusion, using linkage analysis and WGS-based variant analysis, we examined the genetic etiology of ASD and ADHD in the NJLAGS cohort. The family-based design of the cohort allows us to leverage the power of variant segregation while taking into account the phenotype heterogeneity of ASD and ADHD. By examining both small variants and large SVs, we identified genes and pathways, previously known or novel, that could contribute to ASD or ADHD etiology. Because of the difficulty of distinguishing some of the ASD and ADHD behavioral characteristics, understanding the genetic risk factors for the disorders can provide guidance to clinical intervention approaches, both behaviorally and pharmacologically. Future studies, especially functional assessment of these candidate genes, will elucidate their roles in ASD and ADHD and improve the diagnosis and treatment of these disorders.

Acknowledgments

We thank the patients who participated in and contributed to this study. We gratefully acknowledge access to the HPC facilities and support of the computational STEM and bioinformatics scientists from the Office of Advanced Research Computing at Rutgers University. This study was supported by NIMH grants R01 MH070366 and RC1 MH088288 to LMB; New Jersey Governor's Council for Medical Research and Treatment of Autism grant CAUT12APS006 to LMB and CAUT19APL028 to JX and LMB.

Conflicts of Interest

The authors declare no conflict of interest.

Supplementary information

Supplementary information is available at MP's website

Reference

1. Polderman TJ, Hoekstra RA, Posthuma D, Larsson H. The co-occurrence of autistic and ADHD dimensions in adults: an etiological study in 17,770 twins. *Transl Psychiatry*. 2014;4:e435.
2. Rommelse NN, Franke B, Geurts HM, Hartman CA, Buitelaar JK. Shared heritability of attention-deficit/hyperactivity disorder and autism spectrum disorder. *European child & adolescent psychiatry*. 2010;19(3):281-95.
3. Doherty JL, Owen MJ. Genomic insights into the overlap between psychiatric disorders: implications for research and clinical practice. *Genome medicine*. 2014;6.
4. Ghirardi L, Brikell I, Kuja-Halkola R, Freitag CM, Franke B, Asherson P, et al. The familial co-aggregation of ASD and ADHD: a register-based cohort study. *Mol Psychiatry*. 2018;23(2):257-62.
5. Johnson MH, Gliga T, Jones E, Charman T. Annual research review: Infant development, autism, and ADHD--early pathways to emerging disorders. *J Child Psychol Psychiatry*. 2015;56(3):228-47.
6. Ronald A, Hoekstra RA. Autism Spectrum Disorders and Autistic Traits: A Decade of New Twin Studies. *Am J Med Genet B*. 2011;156b(3):255-74.
7. Stergiakouli E, Davey Smith G, Martin J, Skuse DH, Viechtbauer W, Ring SM, et al. Shared genetic influences between dimensional ASD and ADHD symptoms during child and adolescent development. *Mol Autism*. 2017;8:18.
8. Elia J, Gai X, Xie HM, Perin JC, Geiger E, Glessner JT, et al. Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes (vol 15, pg 637, 2010). *Mol Psychiatry*. 2010;15(11):1122-.

9. Martin J, Cooper M, Hamshere ML, Pocklington A, Scherer SW, Kent L, et al. Biological overlap of attention-deficit/hyperactivity disorder and autism spectrum disorder: evidence from copy number variants. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2014;53(7):761-70 e26.
10. McCarthy DJ, Humburg P, Kanapin A, Rivas MA, Gaulton K, Cazier JB, et al. Choice of transcripts and software has a large effect on variant annotation. *Genome medicine*. 2014;6.
11. Bartlett CW, Hou L, Flax JF, Hare A, Cheong SY, Fermano Z, et al. A genome scan for loci shared by autism spectrum disorder and language impairment. *The American journal of psychiatry*. 2014;171(1):72-81.
12. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders, Fifth Edition (DSM-5)*. Washington DC: American Psychiatric Association; 2013.
13. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006;2(12):e190.
14. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nature genetics*. 2002;30(1):97-101.
15. Vieland VJ, Huang Y, Seok SC, Burian J, Catalyurek U, O'Connell J, et al. KELVIN: a software package for rigorous measurement of statistical evidence in human genetics. *Human heredity*. 2011;72(4):276-88.
16. Govil M, Vieland V. Practical Considerations for Dividing Data into Subsets Prior to PPL Analysis. *Hum Hered*. 662008. p. 223-37.

17. Bartlett CW, Flax JF, Logue MW, Vieland VJ, Bassett AS, Tallal P, et al. A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet.* 2002;71(1):45-55.
18. Logue MW, Vieland VJ, Goedken RJ, Crowe RR. Bayesian analysis of a previously published genome screen for panic disorder reveals new and compelling evidence for linkage to chromosome 7. *Am J Med Genet B Neuropsychiatr Genet.* 2003;121b(1):95-9.
19. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-60.
20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-9.
21. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics.* 2011;43(5):491-8.
22. Wang S, Xing J. A primer for disease gene prioritization using next-generation sequencing data. *Genomics & informatics.* 2013;11(4):191-9.
23. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2,504 human genomes. *Nature.* 2015;526(7571):75-81.
24. Ho SS, Urban AE, Mills RE. Structural variation in the sequencing era. *Nat Rev Genet.* 2019;21(3):171-89.

25. Mohiyuddin M, Mu JC, Li J, Bani Asadi N, Gerstein MB, Abyzov A, et al. MetaSV: an accurate and integrative structural-variant caller for next generation sequencing. *Bioinformatics*. 2015;31(16):2741-4.
26. Abyzov A, Li S, Kim DR, Mohiyuddin M, Stütz AM, Parrish NF, et al. Analysis of deletion breakpoints from 1,092 humans reveals details of mutation mechanisms. *Nat Commun*. 2015;6(1):1-12.
27. Chen K, Wallis J, McLellan M, Larson D, Kalicki J, Pohl C, et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nature methods*. 2009;6(9):677-81.
28. Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res*. 2011;21(6):974-84.
29. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*. 2009;25(21):2865-71.
30. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19(5):455-77.
31. Abyzov A, Gerstein M. AGE: defining breakpoints of genomic structural variants at single-nucleotide resolution, through optimal alignments with gap excision. *Bioinformatics*. 2011;27(5):595-603.
32. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156-8.

33. Sedlazeck FJ, Dhroso A, Bodian DL, Paschall J, Hermes F, Zook JM. Tools for annotation and comparison of structural variation. *F1000Res*. 2017;6.
34. Hu H, Huff CD, Moore B, Flygare S, Reese MG, Yandell M. VAAST 2.0: improved variant classification and disease-gene identification using a conservation-controlled amino acid substitution matrix. *Genetic epidemiology*. 2013;37(6):622-34.
35. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020;369(6509):1318-30.
36. Geoffroy V, Herenger Y, Kress A, Stoetzel C, Piton A, Dollfus H, et al. AnnotSV: an integrated tool for structural variations annotation. *Bioinformatics*. 2018;34(20):3572-4.
37. Hu H, Roach JC, Coon H, Guthery SL, Voelkerding KV, Margraf RL, et al. A unified test of linkage analysis and rare-variant association for analysis of pedigree sequence data. *Nature biotechnology*. 2014;32(7):663-9.
38. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-43.
39. Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A, et al. Transcriptional landscape of the prenatal human brain. *Nature*. 2014;508(7495):199-+.
40. Lindsay SJ, Xu YB, Lisgo SN, Harkin LF, Copp AJ, Gerrelli D, et al. HDBR Expression: A Unique Resource for Global and Individual Gene Expression Studies during Early Human Brain Development. *Front Neuroanat*. 2016;10.

41. Cao X, Zhang Y, Abdulkadir M, Deng L, Fernandez TV, Garcia-Delgar B, et al. Whole-exome sequencing identifies genes associated with Tourette's disorder in multiplex families. *Mol Psychiatry*. 2021;26(11):6937-51.
42. Greene CS, Krishnan A, Wong AK, Ricciotti E, Zelaya RA, Himmelstein DS, et al. Understanding multicellular function and disease with human tissue-specific networks. *Nature genetics*. 2015;47(6):569-76.
43. Dickinson ME, Flenniken AM, Ji X, Teboul L, Wong MD, White JK, et al. High-throughput discovery of novel developmental phenotypes. *Nature*. 2016;537(7621):508-+.
44. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607-D13.
45. Kamburov A, Pentchev K, Galicka H, Wierling C, Lehrach H, Herwig R. ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic acids research*. 2011;39(Database issue):D712-7.
46. Sun S, Miller M, Wang Y, Tyc KM, Cao X, Scott RT, Jr., et al. Predicting embryonic aneuploidy rate in IVF patients using whole-exome sequencing. *Human genetics*. 2022.
47. Stoleran ES, Francisco E, Stallworth JL, Jones JR, Monaghan KG, Keller-Ramey J, et al. Genetic variants in the KDM6B gene are associated with neurodevelopmental delays and dysmorphic features. *Am J Med Genet A*. 2019;179(7):1276-86.

48. Insa Pineda I, Gómez González CL. The KDM6B mutation: Phenotype and clinical characteristics-Report of a case. *Rev Psiquiatr Salud Ment.* 2021.
49. Kafle OP, Wang X, Cheng S, Ding M, Li P, Cheng B, et al. Genetic Correlation Analysis and Transcriptome-wide Association Study Suggest the Overlapped Genetic Mechanism between Gout and Attention-deficit Hyperactivity Disorder. *Can J Psychiatry.* 2021;66(12):1077-84.
50. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, et al. Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell.* 2020;180(3):568-84.e23.
51. Li J, Kim S, Kobayashi T, Liang FX, Korzeniewski N, Duensing S, et al. Neurl4, a novel daughter centriole protein, prevents formation of ectopic microtubule organizing centres. *EMBO Rep.* 2012;13(6):547-53.
52. Chang Q, Yang H, Wang M, Wei H, Hu F. Role of Microtubule-Associated Protein in Autism Spectrum Disorder. *Neurosci Bull.* 2018;34(6):1119-26.
53. Ming X, Walters AS. Autism Spectrum Disorders, Attention Deficit/Hyperactivity Disorder, and Sleep Disorders. *Curr Opin Pulm Med.* 2009;15(6).
54. Yang Z, Matsumoto A, Nakayama K, Jimbo EF, Kojima K, Nagata K, et al. Circadian-relevant Genes Are Highly Polymorphic in Autism Spectrum Disorder Patients. *Brain Dev.* 2016;38(1).
55. Satterstrom FK, Walters RK, Singh T, Wigdor EM, Lescai F, Demontis D, et al. Autism spectrum disorder and attention deficit hyperactivity disorder have a similar burden of rare protein-truncating variants. *Nat Neurosci.* 2019;22(12):1961-5.

56. Dong D, Zielke HR, Yeh D, Yang P. Cellular stress and apoptosis contribute to the pathogenesis of autism spectrum disorder. *Autism Res.* 2018;11(7):1076-90.
57. Vergult S, Dheedene A, Meurs A, Faes F, Isidor B, Janssens S, et al. Genomic aberrations of the CACNA2D1 gene in three patients with epilepsy and intellectual disability. *Eur J Hum Genet.* 2015;23(5):628-32.
58. Prasad A, Merico D, Thiruvahindrapuram B, Wei J, Lionel AC, Sato D, et al. A discovery resource of rare copy number variations in individuals with autism spectrum disorder. *G3.* 2012;2(12):1665-85.
59. Smith M, Flodman PL, Gargus JJ, Simon MT, Verrell K, Haas R, et al. Mitochondrial and ion channel gene alterations in autism. *Biochim Biophys Acta.* 2012;1817(10):1796-802.
60. Rong Y, Yang C-J, Jin Y, Wang Y. Prevalence of attention-deficit/hyperactivity disorder in individuals with autism spectrum disorder: A meta-analysis. *Res Autism Spectr Disord.* 2021;83:101759.
61. Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, et al. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Molecular autism.* 2013;4.
62. Schork AJ, Won H, Appadurai V, Nudel R, Gandal M, Delaneau O, et al. A genome-wide association study of shared risk across psychiatric disorders implicates gene regulation during fetal neurodevelopment. *Nat Neurosci.* 2019;22(3):353-61.
63. Zhang L, Chang S, Li Z, Zhang K, Du Y, Ott J, et al. ADHDgene: a genetic database for attention deficit hyperactivity disorder. *Nucleic acids research.* 2012;40(Database issue):D1003-9.

64. Iakoucheva LM, Muotri AR, Sebat J. Getting to the Cores of Autism. *Cell*. 2019;178(6):1287-98.
65. Myers SM, Challman TD, Bernier R, Bourgeron T, Chung WK, Constantino JN, et al. Insufficient Evidence for "Autism-Specific" Genes. *Am J Hum Genet*. 2020;106(5):587-95.
66. Charrier A, Olliac B, Roubertoux P, Tordjman S. Clock Genes and Altered Sleep-Wake Rhythms: Their Role in the Development of Psychiatric Disorders. *Int J Mol Sci*. 2017;18(5).
67. Novikova G, Andrews SJ, Renton AE, Marcora E. Beyond association: successes and challenges in linking non-coding genetic variation to functional consequences that modulate Alzheimer's disease risk. *Mol Neurodegener*. 2021;16(1):27.

Figure legend

Figure 1. Project overview and overall analysis flow.

Figure 2. Linkage analysis for ASD and ADHD. Sequential and pooled analysis results are shown in blue and red lines, respectively. The red dotted line indicates the PPL=0.26, which corresponds to $p=0.01$.

Figure 3. Protein-protein interaction (PPI) networks of candidate genes. Candidate genes were ADHD risk genes in **(A, B)** and ASD/ADHD risk genes in **(C, D)**. **(A, C)**: PPI networks of the candidate genes. Other known NDD genes were added as intermediate nodes if they interact with more than one candidate gene supported by PPI evidence from ≥ 2 databases. **(B, D)**: PPI network of candidate genes that are not in **(A)** and **(C)**, respectively. Known genes were added as intermediate nodes if they interact with more than one candidate gene with PPI evidence from any PPI database. PPI networks were defined by three databases, ConsensusPathDB, STRING, and GIANT_v2. Genes were colored by gene lists. ASD/ADHD genes were used to color genes in **(A, B)** and ADHD genes were used to color genes in **(C, D)**. To simplify the network, PPIs between non-candidate genes were removed. Genes belonging to two GO terms, methylation and cation channel activity, were highlighted in **A, C**, respectively.

Tables

Table 1. Summary of families and their phenotypes.

	All	Male	Female	Male/Female ratio	Families
ASD	98	77	21	3.67	73
ADHD	68	46	22	2.09	47
ASD and ADHD	41	30	11	2.73	36
ASD or ADHD	125	93	32	2.91	73
All samples	493	281	212	1.33	73

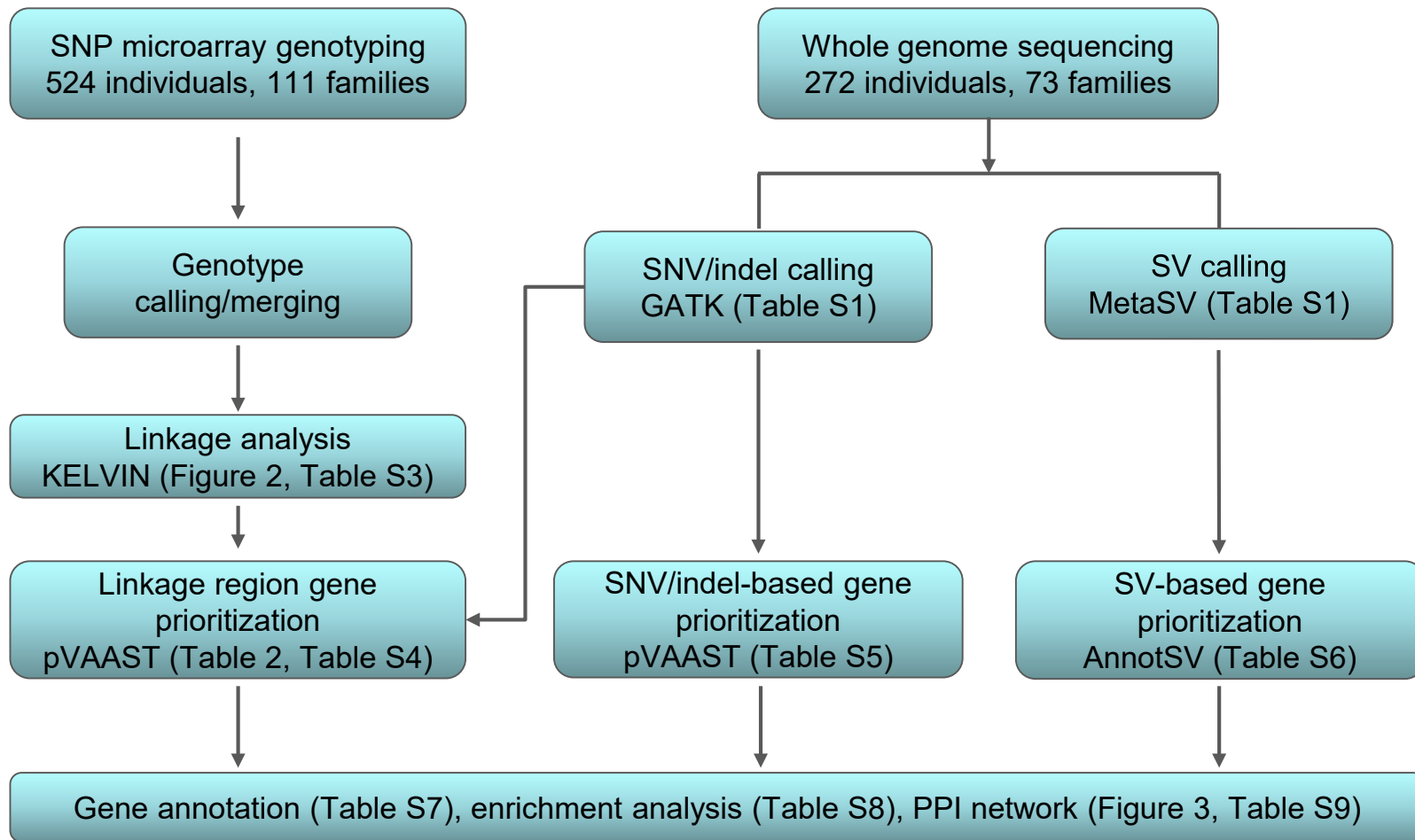
Table 2. Candidate genes in linkage regions

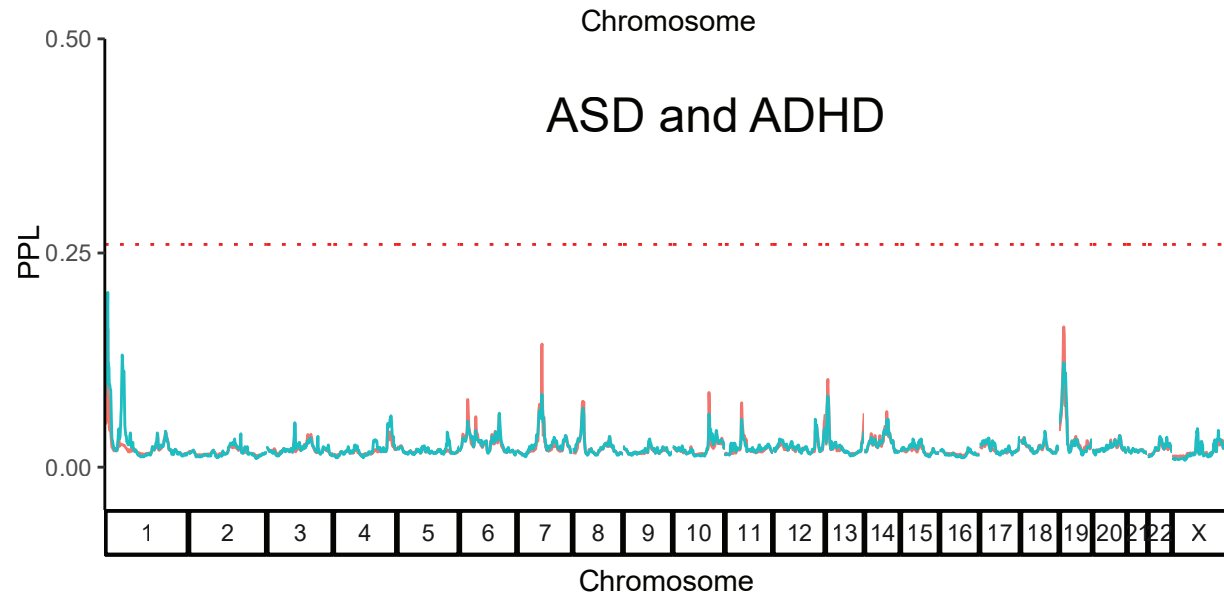
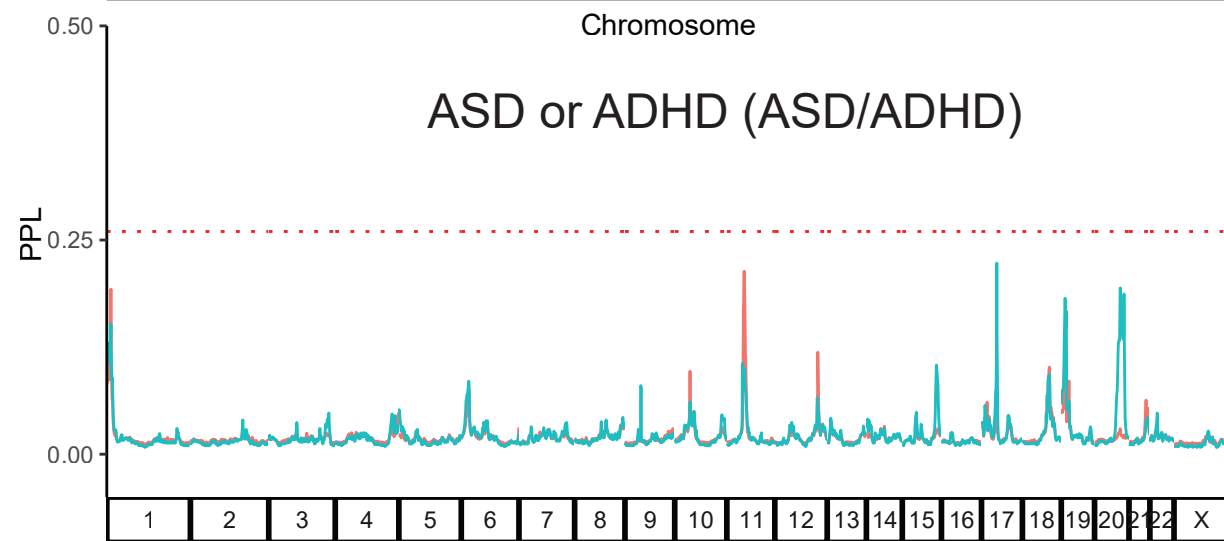
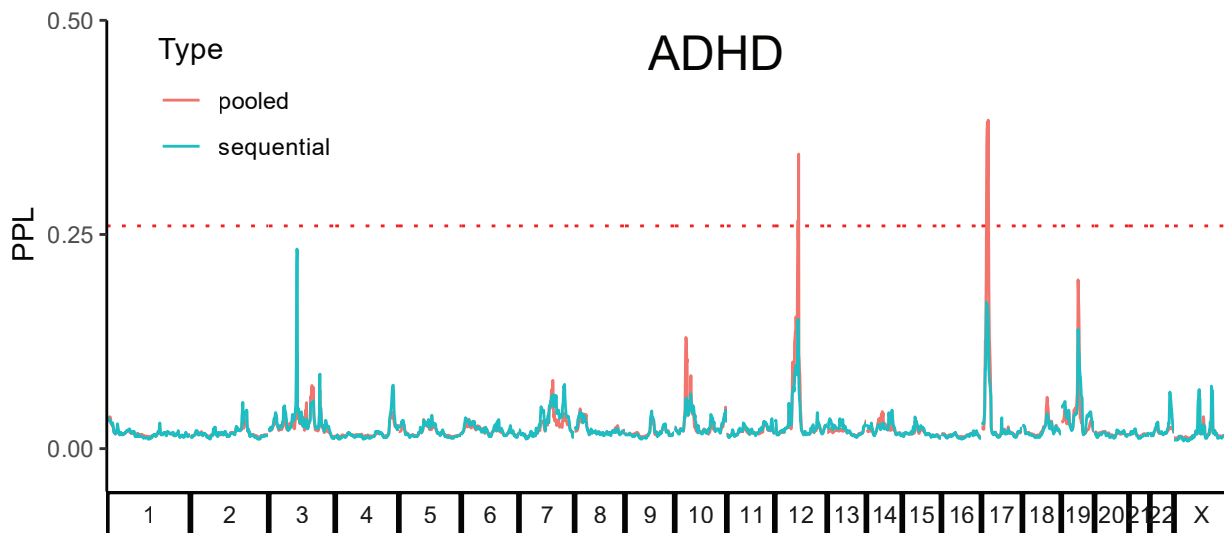
Gene	Chr	Inheritance	Pvalue	pLI	Database
<i>KDM6B</i>	chr17	Dominant	5.63E-04	1	SFARI, iPSYCH
<i>MYH13</i>	chr17	Dominant	5.63E-04	0	
<i>USP6</i>	chr17	Dominant	5.63E-04	0	
<i>MYBBP1A</i>	chr17	Dominant	5.63E-04	0	
<i>KMT2D</i>	chr12	Dominant	1.30E-03	1	
<i>MYO1A</i>	chr12	Dominant	1.30E-03	0	
<i>COL2A1</i>	chr12	Dominant	1.30E-03	1	iPSYCH
<i>PTPRB</i>	chr12	Dominant	1.30E-03	0.94	SFARI
<i>ITGA7</i>	chr12	Dominant	1.30E-03	0	
<i>STAT2</i>	chr12	Dominant	1.30E-03	0.04	
<i>NEURL4</i>	chr17	Dominant	5.63E-04	1	
<i>TIMELESS</i>	chr12	Dominant	1.30E-03	0	
<i>CCDC65</i>	chr12	Dominant	1.30E-03	0	
<i>PFAS</i>	chr17	Dominant	5.63E-04	0	
<i>NLRP1</i>	chr17	Dominant	5.63E-04	0	
<i>EIF4B</i>	chr12	Dominant	1.30E-03	1	
<i>NELL2</i>	chr12	Dominant	1.30E-03	1	
<i>WNT10B</i>	chr12	Dominant	1.30E-03	0	
<i>ALG10B</i>	chr12	Dominant	2.60E-03	0	
<i>PELP1</i>	chr17	Dominant	1.69E-03	1	
<i>CHRNE</i>	chr17	Dominant	1.69E-03	0	
<i>GRASP</i>	chr12	Dominant	3.90E-03	0.93	
<i>ARHGEF15</i>	chr17	Dominant	2.82E-03	0	
<i>TROAP</i>	chr12	Dominant	6.50E-03	0	
<i>TRPV3</i>	chr17	Dominant	4.50E-03	0	
<i>P2RX5</i>	chr17	Dominant	4.50E-03	NA	
<i>DNAJC22</i>	chr12	Dominant	1.04E-02	0	
<i>ACADVL</i>	chr17	Dominant	1.13E-02	0	
<i>ZZEF1</i>	chr17	Dominant	1.13E-02	1	
<i>TEKT1</i>	chr17	Dominant	1.13E-02	0	
<i>ANKRD33</i>	chr12	Dominant	2.60E-02	1	
<i>PER1</i>	chr17	Dominant	1.13E-02	0.88	SFARI, ADHDgene
<i>LYZ</i>	chr12	Dominant	2.60E-02	0	
<i>USP43</i>	chr17	Dominant	1.69E-02	0.01	
<i>PELP1</i>	chr17	Recessive	5.63E-04	1	
<i>SHPK</i>	chr17	Recessive	1.13E-03	0	
<i>ESPL1</i>	chr12	Recessive	3.90E-03	1	

Table 3. Candidate gene summary

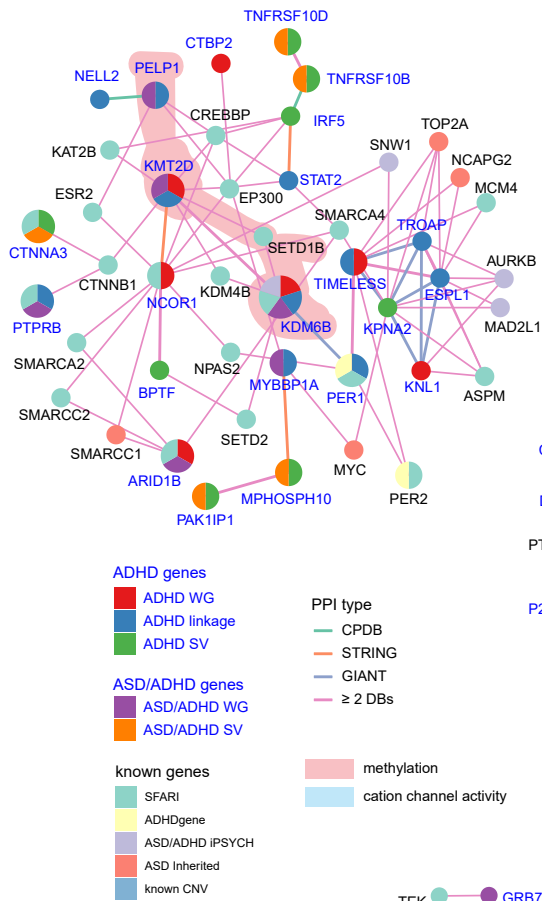
	SNV/Indel		SV		Total (unique)	
	Genes	Families*	Genes	Families*	Genes	Families*
ADHD						
Linkage	36	19	0	0	36	19
WGS	40	47	43	42	83	47
Total (Unique)	70	47	43	42	113	47
ASD/ADHD						
WGS	94	73	44	73	138	73
Total (Unique)	140	73	61	73	207	73

*For SNV/indel, families with segregating variants in a given gene were counted. For SV, families with at least one affected individual with an exonic SV were counted

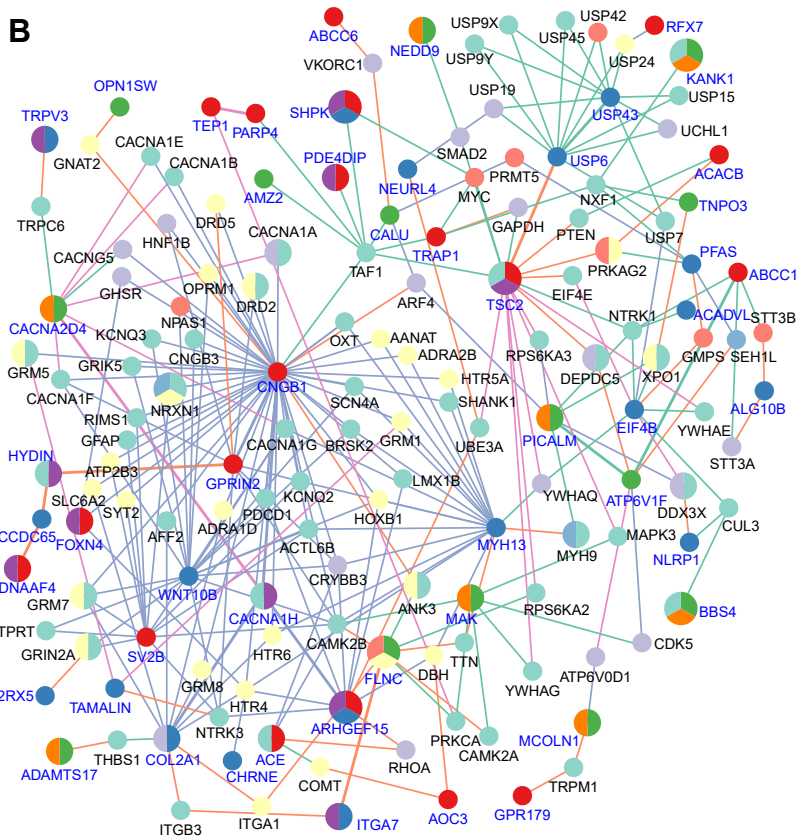




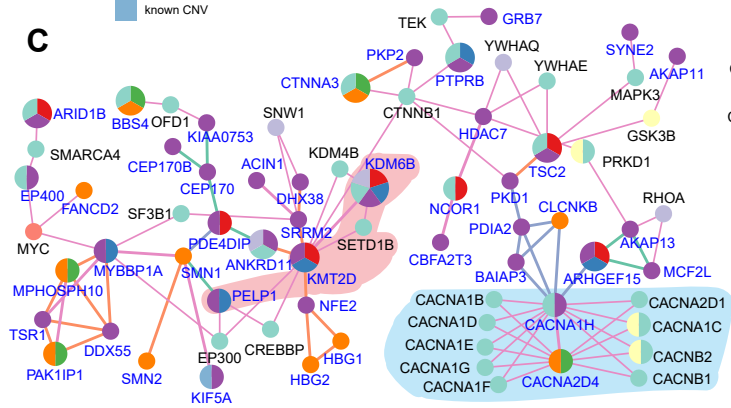
A



B



C



D

