

# SOX7: Novel Autistic Gene Identified by Analysis of Multi-Omics Data

**Samantha Gonzales**

Florida International University

**Jane Zizhen Zhao**

Miami Dade College Kendall Campus and School for Advanced Studies

**Na Young Choi**

University of North Texas

**Prabha Acharya**

University of North Texas

**Sehoon Jeong**

Inje University

**Moo-Yeal Lee** (✉ [Moo-Yeal.Lee@unt.edu](mailto:Moo-Yeal.Lee@unt.edu))

University of North Texas

---

## Research Article

**Keywords:** Genome-wide association studies (GWAS), autism spectrum disorder (ASD), gene-based association test, RNA-seq data analysis

**Posted Date:** September 14th, 2023

**DOI:** <https://doi.org/10.21203/rs.3.rs-3346245/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

## Background

Despite thousands of variants identified by genome-wide association studies (GWAS) to be associated with autism spectrum disorder (ASD), it is unclear which mutations are causal because most are noncoding. Consequently, reliable diagnostic biomarkers are lacking. RNA-seq analysis captures biomolecular complexity that GWAS cannot by considering transcriptomic patterns. Therefore, integrating DNA and RNA testing may reveal causal genes and useful biomarkers for ASD.

## Methods

We performed gene-based association studies using an adaptive test method with GWAS summary statistics from two large Psychiatric Genomics Consortium (PGC) datasets (ASD2019: 18,382 cases and 27,969 controls; ASD2017: 6,197 cases and 7,377 controls). We also investigated differential expression for genes identified with the adaptive test using an RNA-seq dataset (GSE30573: 3 cases and 3 controls) and DESeq2.

## Results

We identified 5 genes significantly associated with ASD in ASD2019 (KIZ-AS1,  $p = 8.67 \times 10^{-10}$ ; KIZ,  $p = 1.16 \times 10^{-9}$ ; XRN2,  $p = 7.73 \times 10^{-9}$ ; SOX7,  $p = 2.22 \times 10^{-7}$ ; LOC101929229 (also known as PINX1-DT),  $p = 2.14 \times 10^{-6}$ ). Two of the five genes were replicated in ASD2017: SOX7 ( $p = 0.00087$ ) and LOC101929229 ( $p = 0.009$ ), and KIZ was close to the replication boundary of replication ( $p = 0.06$ ). We identified significant expression differences for SOX7 ( $p = 0.0017$ , adjusted  $p = 0.0085$ ), LOC101929229 ( $p = 5.83 \times 10^{-7}$ , adjusted  $p = 1.18 \times 10^{-5}$ ), and KIZ ( $p = 0.00099$ , adjusted  $p = 0.0055$ ). SOX7 encodes a transcription factor that regulates developmental pathways, alterations in which may contribute to ASD.

## Limitations:

The limitation of the gene-based analysis is the reliance on a reference population for estimating linkage disequilibrium between variants. The similarity of this reference population to the population of study is crucial to the accuracy of many gene-based analyses, including those performed in this study. As a result, the extent of our findings is limited to European populations, as this was our reference of choice. Future work includes a tighter integration of DNA and RNA information as well as extensions to non-European populations that have been under-researched.

## Conclusions

These findings suggest that SOX7 and its related SOX family genes encode transcription factors that are critical to the downregulation of the canonical Wnt/ $\beta$ -catenin signaling pathway, an important developmental signaling pathway, providing credence to the biologic plausibility of the association between gene SOX7 and autism spectrum disorder.

## Background

Autism spectrum disorder (ASD) is a heterogeneous grouping of neurodevelopmental traits that is diagnosed in approximately 1% of the world population (Fombonne, 2009). ASD conditions are characterized by attention-deficit hyperactivity disorder (ADHD), intellectual disability (ID), epilepsy, social communication deficits and restricted, repetitive, or unusual sensory-motor behaviors, or gastrointestinal problems (Gillberg et al., 2014). Extensive research efforts have gone into understanding the causes of individual differences in autistic behavior. Twin and family studies strongly demonstrate that autism has a particularly large genetic basis, with estimated heritability ranging from 40–90% (Devlin et al., 2013; Gaugler et al., 2014; Sandin et al., 2017; Tick et al., 2016). Molecular genetic studies have revealed that the genetic risk for autism is shaped by a combination of rare and common genetic variants (Consortium, 2017).

Over the past decade, genome-wide association studies (GWAS) and other types of genetic studies have identified increasing numbers of single nucleotide polymorphisms (SNPs) (Grove et al., 2019; Robinson et al., 2016) and other forms of genetic variation that are associated with ASD (Bourgeron, 2015). It has been estimated that more than 100 genes and genomic regions are associated with autism (Sanders et al., 2015; Satterstrom et al., 2020). While most of these studies focused on identifying heritable single nucleotide polymorphisms (SNPs) associated with ASD risk, other studies have demonstrated the influence of de novo mutations ranging from a single base (O’Roak et al., 2012; Sanders et al., 2012) thousands to millions of bases long (Levy et al., 2011; Sanders et al., 2011), such as copy number variants (CNVs). Several likely gene-disruptive (LGD) variants in genes such as GRIK2 (Jamain et al., 2002) and ASMT (Melke et al., 2008) affecting autism risk were found exclusively or more frequently in individuals with autism than in control groups. Additional evidence strongly suggests that mutations in NLGN3 and NLGN4 are involved in autism (Jamain et al., 2003). Additionally, deletions at Xp22.3 that include NLGN4 have been reported in several autistic individuals. Roohi et al. found that (Roohi et al., 2009) CNTN4 plays an essential role in the formation, maintenance, and plasticity of neuronal networks. Disruption of CNTN4 is known to cause developmental delay and mental retardation. This report suggests that mutations affecting CNTN4 function may be relevant to ASD pathogenesis. A review by Li and Brown (Li et al., 2012) discussed a substantial body of evidence resulting from genome-wide screening for several widely studied candidate ASD genes. Similarly, a large-scale international collaboration was conducted to combine independent genotyping data to improve statistical power and aid in robust discovery of GWS loci (Consortium, 2017). This international collaboration also identified a significant genetic correlation between schizophrenia and autism with several neurodevelopmental-related genes, such as EXT1, ASTN2, MACROD2, and HDAC4. A combined analysis investigating both rare and common gene variants supported the evidence of the role of several genes/loci associated with

autism (e.g., NRXN1, ADNP, 22q11 deletion) and revealed new variants in known autism-risk genes such as ADNP, NRXN1, NINL, and MECP2 and identified new compelling candidate genes such as KALRN, PLA2G4A, and RIMS4 (Leblond et al., 2019). Recently, Buxbaum (Buxbaum, 2022) summarized the prevalence of some genetic variants in subjects ascertained for ASD.

Research investigating the gene expression profiles of those with ASD has also proven insightful genetic contributions to ASD. The expression levels of genes containing rare mutations associated with autism were evaluated in lymphoblasts from autism cases and controls, including the aforementioned genes, such as NLGN3, NLGN4, NRXN1, and MeCP2. Of these, NLGN3 was found to be differentially expressed along with SHANK3 (Yasuda et al., 2011). More comprehensive gene expression analyses have confirmed susceptibility genes previously reported in GWAS-based analysis and identified novel differentially expressed genes and biological pathways enriched for these genes (Rahman et al., 2020). RNA sequencing data analyses have elucidated several potential drivers of autism susceptibility, such as resting-state functional brain activity (Berto et al., 2022), dopaminergic influences in the dorsal striatum (Brandenburg et al., 2020), overexpression of FOXP1, a gene involved in regulating tissue and cell type-specific gene transcription in the brain (Chien et al., 2013; Ferland et al., 2003), genome-wide alterations to lncRNA levels, downregulation of alternative splicing events, and brain region-dependent alterations in gene expression (Parikshak et al., 2016). These studies indicate that integrating GWAS and RNA-seq data analysis can provide a better picture of the various underlying mechanisms behind a heterogeneous, multifaceted condition such as ASD.

In this study, we performed whole-genome gene-based association tests for ASD with the adaptive test method (Guo & Wu, 2018) using summary statistics from two large GWAS datasets obtained from the Psychiatric Genomics Consortium (PGC). We identified 5 genes significantly associated with ASD in the ASD 2019 data. Among these 5 genes, the genes SOX7 and LOC101929229 (also known as PINX1-DT) were replicated in ASD 2017 data. Gene LOC101929229 is ncRNA. Further RNA sequencing data analysis indicated that the gene SOX7 was significantly upregulated in cases compared to controls. SOX7 encodes a member of the SOX (SRY-related HMG-box) family of transcription factors that pivotally contribute to determining cell fate and identity in many lineages. The encoded protein may function as a transcriptional regulator after forming a protein complex with other proteins, leading to autism.

## Methods

### Datasets

*Discovery GWAS summary statistics:* The discovery dataset (labeled as asd2019) includes summary statistics from a meta-analysis of European samples derived from two cohorts: a population-based case control study from the iPSYCH project and a family trio-based study from the Psychiatric Genomics Consortium (PGC) (Grove et al., 2019). The iPSYCH samples included individuals born by a known mother who was a resident of Denmark at the time of their first birthday. Cases were identified using the Danish Psychiatric Central Research Register, using diagnoses from 2013 or earlier by psychiatrists

according to diagnostic code ICD10, which includes diagnoses of childhood autism, atypical autism, Asperger's syndrome, "other pervasive developmental disorders", and "pervasive developmental disorder, unspecified" (Grove et al., 2019). The PGC samples consisted of 5 cohorts, whose trios were analyzed as cases and pseudo controls. Details regarding these studies can be found in (Grove et al., 2019) and (Consortium, 2017). The combined sample size consisted of 18,382 cases and 27,969 controls. Imputation and quality control were performed *via* PGC's Ricopili pipeline, which ensures the production of robust, reproducible, and comparable datasets. The iPSYCH samples were processed separately in the 23 genotyping batches, while the PGC samples were processed separately for each study. Genotype imputation was performed with IMPUTE2/SHAPEIT (Bryan N. Howie, 2009; Delaneau et al., 2011) in the Ricopili pipeline using the 1000 Genomes Project phase 3 dataset as the reference set. Regions demonstrating high linkage disequilibrium were excluded, and one of the highly similar pairs of subjects identified by PLINK's identity by state (IBS) analysis (Chang et al., 2015) was reduced at random, with a preference for retaining cases. Association was performed using PLINK on imputed dosage data, and the meta-analysis was performed using METAL (Grove et al., 2019). More detailed descriptions of each stage of the analysis can be found in Grove et al. (Grove et al., 2019). The summary statistics produced by this study and subsequently used for our analysis can be found at <https://pgc.unc.edu/for-researchers/download-results/>.

*Replication GWAS summary statistics:* The replication dataset (labeled as asd2017) includes summary statistics from a European-ancestry meta-analysis performed by the Autism Spectrum Disorders Working Group (AWG) of The Psychiatric Genomics Consortium (PGC), which aimed at improving statistical power to detect loci significantly associated with ASD. The meta-analysis was performed on data from 14 independent cohorts across different ancestries totaling over 16,000 individuals. For each step in the meta-analysis, each cohort was processed individually. Individuals were excluded if they were assessed at less than 36 months of age or if diagnostic criteria were not met from the Autism Diagnostic Interview-Revised (ADI-R) or the Autism Diagnostic Observation Schedule (ADOS) domain scores. While a "worldwide" meta-analysis on this aggregate dataset was performed, we derive our replication dataset based on the smaller European-only analysis consisting of 6,197 ASD cases and 7,377 controls (Consortium, 2017). Each stage of the imputation and quality control was performed similarly to the asd2019 data: Imputation and quality control on PGC samples were performed following the PGC's "Ricopili" pipeline. Since multiple studies were involved, necessary studies were performed to check for and remove duplicate individuals prior to imputation. Family trio-based data were organized as case and pseudo controls. Criteria for SNP retention and other pre-imputation quality control steps can be found in the study's supplementary File 1 (Consortium, 2017). Genotype imputation was performed with IMPUTE2/SHAPEIT using the 2,184 phased haplotypes from the full 1000 Genomes Project dataset as the reference set. All 14 cohorts were tested for associations individually using an additive logistic regression model in PLINK. More detailed information about each stage of the analyses performed by this study can be found in supplementary File 1 (Consortium, 2017). The resulting summary statistics that were utilized in our analysis can be found at <https://pgc.unc.edu/for-researchers/download-results/>.

## Bulk RNA-Seq

The RNA dataset was obtained from a gene coexpression analysis that aimed to identify modules of coexpressed genes associated with ASD (Voineagu et al., 2011). The study can be found in the Gene Expression Omnibus (GEO) database under accession number GSE30573. Detailed descriptions of the raw data acquisition and quality control processes can be found in the supplementary information of (Voineagu et al., 2011) as well as the GEO accession viewer. Briefly, brain tissue samples (frontal cortex, temporal cortex, and cerebellum) were obtained from the Autism Tissue Project (ATP) and the Harvard Brain Bank. Cases were diagnosed using ADI-R diagnostic scores, which can be found along with other clinical data upon request from the ATP website. Total RNA was extracted from the sample tissues following the Qiagen miRNA kit instructions. Quality and concentration were assessed by an Agilent Bioanalyzer and Nanodrop, respectively. Reads were generated using an Illumina Genome Analyzer II sequencer using the manufacturer settings and were 73–76 nucleotides in length. Raw sequencing data for the frontal and temporal cortex samples were available in the SRA run selector for 3 autism cases and 3 controls (Voineagu et al., 2011).

## Quality Control & Preprocessing

### GWAS summary data

After downloading the raw summary statistics from the PGC website, we performed quality control analysis to ensure robust and high-quality results. Only SNPs on autosomal chromosomes were used. First, SNPs with an imputation information metric (INFO) score  $> 0.9$  were removed. Next, SNPs with strand-ambiguous alleles or nonbiallelic loci were removed as well as SNPs with duplicate rs IDs. Z scores were then calculated using each variant's odds ratio and standard error using the equation. After quality control, the raw variants were sorted into hg19 Reference Sequence genes. Linkage disequilibrium (LD) within each gene was calculated using the 1000 Genomes European reference panel (phase 3). For each gene, a subset of the GWAS variants of the gene's transcription start site and transcription end site were matched to the reference variants, ensuring that both used the same reference allele and flipping Z score signs if necessary. Genes that contained less than 2 SNPs were removed. Pearson's correlation between this subset of genotypes was calculated and used as the gene-wide LD. One SNP of a pair of SNPs with perfect correlation (within a gene) was removed. The processed data were saved in 22 'RData' files (one for each chromosome) containing a list of data frames, where each list element comprised 1) SNP information for a specific gene and 2) its corresponding LD matrix.

### RNA-seq data

The sequence read archive (SRA) accession list and associated sample metadata ("SRA Run Table") for GSE30573 were downloaded from the SRA run selector page for the study. Raw fastq files were downloaded from the SRA using the SRA Toolkit *via* the 'prefetch' and 'fastq-dump' commands ("SRA Toolkit"). We used FastQC to assess the quality of reads in each file and MultiQC to visualize the results in batch format (Andrews, 2010; Ewels et al., 2016). Only 1 sample failed the 'per sequence base quality' assessment and was subsequently trimmed of low-quality reads using the command-line tool

'fastq\_quality\_filter' from the FastX toolkit using a minimum quality Phred score of 20 and a minimum percent of bases per read to meet that threshold of 50% ("FASTX-Toolkit,"). Reassessment *via* FastQC demonstrated this as sufficient trimming to meet the quality needed for downstream analysis.

## Statistical Analysis

*Gene-based Association Test.* To perform gene-based association testing, we used the function 'sats' in the R package 'mkatr' (Guo & Wu, 2018). This function computes p values for 3 different SNP-set testing methods using GWAS summary statistics and an LD matrix calculated from a reference panel. A brief description of each is as follows: Let  $m$  denote the number of variants considered in a gene or gene region and let  $(z_1, \dots, z_m)$  represent the GWAS summary statistics for each gene. Let  $R = (r_{ij})$  denote the estimated correlation between Z statistics based on variant linkage disequilibrium (LD) calculated from a reference panel (Guo & Wu, 2018). The tests included in the sats function are the sum test (a type of burden test), the squared sum test and the adaptive test. The three tests are as follows:

1. Sum test (ST):  $B = \sum_{j=1}^m z_j$
2. Squared sum test (S2T):  $Q = \sum_{j=1}^m z_j^2$
3. Adaptive test (AT):  $T = \min_{\rho \in [0,1]} P(Q_\rho)$ , where  $Q_\rho = (1 - \rho)Q + \rho B^2$  and  $P(Q_\rho)$  denotes the p value of  $Q_\rho$ .

It can be shown that  $Q_\rho$  asymptotically follows a weighted sum of independent chi-squared distribution with 1 degree of freedom [ $\chi^2(df = 1)$ ] whose weights equal the eigenvalues of  $R$ . The minimum p value of AT is searched for over a range of  $\rho$  in the interval  $[0, 1]$  (Guo & Wu, 2018).

The ST is most valuable when all variants have the same direction of effect and approximately equal effect size, while the S2T will perform better than ST when variants have different directions of effects. AT utilizes information from both ST and S2T, meaning AT can adapt to the variants in the data better than ST or S2T alone. Indeed, the adaptive test shows the most robust performance across a wider range of scenarios (Guo & Wu, 2018). In this study, we report results based on the AT method. More details regarding the derivation of these tests and their relation to the single-variant association test can be found in (Guo & Wu, 2018).

*Differential Gene Expression Analysis.* After passing quality control, RNA-seq reads were aligned to a reference genome using STAR (Dobin et al., 2013) by the following two steps: genome indexing and alignment to the indexed reference genome. We generated the genome index files using STAR's –genomeGenerate flag and setting the option –sjdbOverhang to 75 to match the maximum read length across the samples. The reference genome FASTA file and corresponding annotation GTF file (GChr37/hg19, release 41) used to generate these index files were downloaded from GENCODE (Frankish et al., 2019). After alignment, we used HTSeq (Putri et al., 2022) to estimate the number of reads per gene region. For genes with gene expression counts of at least 10, we used the R package DESeq2 (Love et al.,

2014) to perform differential expression analysis based on normalized gene expression counts. DESeq2 uses a generalized linear model to model the relationship between a trait and the log2-fold changes in gene expression (Love et al., 2014). We used the adjusted p value to assess significance in gene expression differential analysis.

**Computing Environment**

RNA-seq quality control, alignment, and counts were processed on the lonestar6 high-performance cluster provided by TACC at the University of Texas at Austin. Differential gene expression analysis and gene-based association tests were performed in a local Linux (Windows Linux Subsystem) environment using R in RStudio (RStudio 2022.07.1 + 554).

**Results**

**Gene-based Association Test**

**Discovery GWAS**

Out of approximately 19,000 genes tested for association with ASD, 5 genes were identified as significant with Bonferroni corrected p values less than (Fig. 1 and Table 1). SOX7 () encodes a transcription factor involved in regulating embryonic development and cell fate determination (Takash et al., 2001). KIZ () encodes “Kizuna centrosomal protein”, which plays a central role in stabilizing the pericentriolar region before the spindle formation step in cellular division (Oshimori et al., 2006). A gene region that encodes a long noncoding antisense RNA for KIZ, KIZ-AS1 (), was also identified as significant; however, the function of this antisense RNA has not been determined. XRN2 () encodes a 5’-3’ exoribonuclease that is pertinent in promoting transcriptional termination (Eaton & West, 2018). Finally, LOC101929229 (), also known as PINX1-DT, is a lncRNA that is considered a “divergent transcript” of the protein coding gene PINX1. While the divergent transcript function is not defined, PINX1 encodes a protein that enables telomerase RNA binding and inhibitor activity and is involved in several related processes, including DNA biosynthesis and protein localization (Johnson, 2011).

**Replication GWAS**

Among these 5 genes, the genes SOX7 (=0.00087), LOC101929229 (=0.009), and KIZ-AS1 (=0.059) were replicated in the ASD 2017 data. KIZ (=0.06) was close to the boundary of replication in the ASD 2017 data (Table 1).

**Table 1.** Genetic features of significant genes identified in the discovery dataset.



Gene	Chr.	Discovery (asd2019)	Replication (asd2017)	RNA-seq (GSE30573)		
		p value	p value	LFC	p value	Adj. p value
<i>SOX7</i>	8	<b>2.22E-07</b>	<b>0.0009</b>	1.17	<b>0.0017</b>	<b>0.0085</b>
<i>LOC101929229</i>	8	<b>2.14E-06</b>	<b>0.009</b>	3.22	<b>5.83x10<sup>-7</sup></b>	<b>1.18x10<sup>-5</sup></b>
<i>XRN2</i>	20	7.73E-09	0.10	0.34	0.001	0.007
<i>KIZ</i>	20	1.16E-09	0.06	0.63	0.001	0.006
<i>KIZ-AS1</i>	20	8.67E-10	0.059	0.062	0.43	0.58

Abbreviations: Chr.: chromosome; LFC: log2-fold change.

LOC101929229 is also known as PINX1-DT, which is a ncRNA gene. Significance level= $2.5 \times 10^{-6}$  in the analysis with the discovery dataset; 0.05 in the analysis with the replication dataset and the RNA-seq dataset. Bold highlighting indicates significance in all three (discovery, replication, and RNA-seq) analyses.

## Differential Gene Expression Analysis

Among the five genes identified in the discovery of GWAS, the genes SOX7 (log2FoldChange [LFC] = 1.17, ; Benjamini–Hochberg (BH) adjusted  $p = 0.0085$ ), LOC101929229 (LFC=3.22,  $p=5.83 \times 10^{-7}$ , adjusted  $p=1.18 \times 10^{-5}$ ), and KIZ (LFC=0.63,  $p=0.00099$ , BH adjusted  $p=0.0055$ ) were also identified as significant in the differential gene expression analysis (Table 1). A comparison of case–control gene expression counts for SOX7 can be found in Figure 2, demonstrating that SOX7 is consistently upregulated in autism cases compared to controls. The expression of SOX7 is increased in autism patients relative to controls by a multiplicative factor of 2.25. In addition, the expression of LOC101929229 was increased in autism patients compared with controls by a multiplicative factor of 9.31.

## Discussion

Through powerful gene-based analysis, we identified 5 gene regions (KIZ, KIZ-AS1, XRN2, LOC101929229, and SOX7) significantly associated with ASD. Gene SOX7 and LOC101929229 (also known as PINX1-DT) were replicated by a different GWAS data (gene KIZ was close to the boundary of replication ( $p=0.06$ )) and advocated by the differential gene expression analysis performed on publicly available RNA-seq data.

KIZ is located on chromosome 20 and encodes Kizuna centrosomal protein, which aids in stabilizing the pericentriolar region of centrosomes before spindle formation. KIZ has been identified as significantly associated with autism in previous GWAS (Grove et al., 2019), TWAS (Huang et al., 2021), gene-based

analysis (Alonso-Gonzalez et al., 2019), and methylation-based studies (Hannon et al., 2018), and the involvement of cell cycle regulation in autism susceptibility which has also been implicated in previous research (Packer, 2016; Pramparo et al., 2015). KIZ has also been found to be a potentially shared genetic locus between ASD and attention-deficit hyperactivity disorder (ADHD), providing support for its involvement in neurological conditions (Baranova et al., 2022).

XRN2 is located next to KIZ and encodes a 5'-3' exonuclease that is involved in myriad RNA management processes, including transcriptional termination, miRNA expression regulation, nonsense-mediated mRNA decay, and rRNA maturation (Brannan et al., 2012; Nagarajan et al., 2013; Wang & Pestov, 2011; West et al., 2004). XRN2 has been found to play a role in regulating miRNA expression in neurons specifically, and altered miRNA expression regulation has been investigated as a potential mechanism for autism susceptibility (Abu-Elneel et al., 2008; Ghahramani Seno et al., 2011; Hicks & Middleton, 2016; Kinjo et al., 2013; Wu et al., 2016). Likewise, disruption of proper RNA metabolism as a result of altered expression of RNA binding proteins has been implicated in neurological disease as a whole, and the XRN gene family is involved in nonsense-mediated decay of mRNA, a process that has been implicated in autism pathophysiology (Marques et al., 2022; Nussbacher et al., 2019). Previous GWAS have reported SNPs in the region containing XRN2 to be significantly associated with ASD, affirmed by gene-based analysis using MAGMA (Grove et al., 2019). Additionally, a transcriptome-wide association study (TWAS) found XRN2 to be significantly upregulated in autism, in accordance with our findings (Pain et al., 2019). Another gene-based analysis found XRN2 to be associated with ASD, and upon further investigation *via* gene-network analysis and enrichment analysis, not only does XRN2 interact with several genes in the cAMP signaling pathway and RNA transport network, but the enriched KEGG/GO terms for XRN2 (spliceosome, RNA transport, and nucleic acid binding) found to be associated with ASD are also essential processes pivotal to early development (Alonso-Gonzalez et al., 2019). The extensive involvement of XRN2 in such complex mechanisms of gene expression regulation, particularly in neuronal cell types, offers possible insights into the vast heterogeneity of ASD and its overlap with other neurodevelopmental conditions. In fact, more recent research efforts have focused on ascertaining genetic commonalities between ASD and related disorders such as ADHD, obsessive compulsive disorder (OCD), and Tourette syndrome, of which XRN2 seems to be a shared significant locus (Peyre et al., 2020; Yang et al., 2021).

SOX7 is of particular interest due to its hallmark involvement in the regulation of the Wnt/  $\beta$ -catenin pathway, an important developmental signaling pathway. SOX7 and its related SOX family genes encode transcription factors that are critical to the downregulation of the canonical Wnt/  $\beta$ -catenin signaling pathway, which controls embryonic development and adult homeostasis and is involved in a multitude of cellular processes (Kato, 2002; MacDonald et al., 2009). While the Wnt pathway is ubiquitous to nearly all tissue types, proteins involved in Wnt signaling in the brain specifically have been found to localize in the synapses and influence synaptic growth, and knockout murine models of ASD risk genes that are a part of the Wnt pathway have provided support for the disruption of this pathway in autism-like behaviors (Kwan et al., 2016). Indeed, the Wnt/  $\beta$ -catenin signaling pathway has been suggested as a possible avenue for autism pathogenesis in several studies (Caracci et al., 2021; de la Torre-Ubieta et al.,

2016; El Khouri et al., 2021; Hormozdiari et al., 2015; Kwan et al., 2016; Quesnel-Vallières et al., 2019; Vallée et al., 2019).

SOX7 also regulates angiogenesis, vasculogenesis, and endothelial cell development, and the SOX family of transcription factors is critical to cardiovascular development (Francois et al., 2010; Kim et al., 2016). For example, SOX7 was found to be upregulated in sustained hypoxic environments, mediating angiogenesis (Klomp et al., 2020), and a knockout model of SOX7 was found to result in profound vascular defects, demonstrating that SOX7 has an essential role in vasculogenesis and angiogenesis in early development (Lilly et al., 2017). Links between the role of SOX7 in developmental delay and congenital heart disease have been investigated. Specifically, deletions in the region where SOX7 resides have been demonstrated to simultaneously cause congenital heart defects and intellectual disability (Páez et al., 2008; Wat et al., 2009).

Additionally, Wnt signaling has been demonstrated to orchestrate the differentiation of neural vasculature, such as the blood–brain barrier (Reis & Liebner, 2013; Stenman et al., 2008). Likewise, there is evidence of vascular involvement in the development of autism (Casanova, 2007; Emanuele et al., 2010; Ouellette et al., 2020; Yao et al., 2006). One review in particular suggests that mutations affecting the delicate interactions between Wnt signaling and Shh pathways may alter blood brain barrier integrity in autism by aberrantly interacting with neurovascular molecules (Gozal et al., 2021).

Last, oxidative stress has been researched as a potential source of autism susceptibility (Bjørklund et al., 2020; Chauhan & Chauhan, 2006), and the interaction between altered vasculature and autism during oxidative stress could point to another potential source of pathogenesis (Yao et al., 2006). Indeed, the role of Wnt/  $\beta$ -catenin signaling in oxidative stress has been directly implicated in autism susceptibility (Zhang et al., 2012). This combination of evidence that implicates both Wnt signaling and SOX7 interactions in the multitude of interrelated processes that have been suggested as mechanisms behind the etiology of ASD, supplemented by our findings, provides ever-mounting support for more in-depth investigations of these particular genes and pathways.

Wnt/  $\beta$ -catenin, oxidative stress, and impaired/altered vasculature have all been implicated in the development of ASD. These three factors are involved with each other and multiple systemic processes, which may contribute to ASD symptom heterogeneity. The fact that SOX7 is involved in the regulation of Wnt/  $\beta$ -catenin and vasculogenesis points to a potential converging mechanism behind the pathophysiology of ASD. Additionally, the association of SOX7 with autism has been investigated directly. A case study involving a child patient exhibiting “8p23.1 duplication syndrome” revealed a de novo 1.81 Mbp duplication event on chromosome 8 (8p23.1), spanning the region where SOX7 lies (Weber et al., 2014). This patient exhibited characteristic symptoms of the duplication syndrome, including delay of motor and speech development and intellectual disability, which heavily overlap with autism and related intellectual conditions. Indeed, this patient also exhibited symptoms specific to ASD, such as repetitive compulsive behavior.

A GWAS performed in a Mexican population found that SOX7 was differentially methylated between autism cases and controls (Aspra et al., 2022). Another study also found that differential methylation was associated with an “elevated polygenic burden” for autism and further identified that two significantly associated CpG sites were located near GWAS markers for autism on chromosome 8 in the same region as SOX7 (Hannon et al., 2018). It is worth noting that this study also found evidence of SNPs associated with both autism and DNA methylation that were annotated to KIZ and XRN2, two genes that we also found to be significantly associated with ASD.

Changes in methylation lead to changes in gene expression, providing another plausible mechanism of SOX7 involvement: a change in SOX7 methylation affects the expression and thus availability of the transcription factor it encodes, which has a downstream effect on the subsequent pathways SOX7 regulates, such as Wnt/  $\beta$ -catenin. Indeed, both methylation studies demonstrated a negative difference in methylation between autism cases and controls. Generally, undermethylation results in a less compact 3-dimensional genome structure, allowing for greater access to the gene and an increase in expression, which we see in the higher gene expression counts in autism cases versus controls in our RNA-seq data (Figure 2) (Buitrago et al., 2021; Keshet et al., 1986; Lewis & Bird, 1991).

Finally, altered expression of SOX7 has been shown to play a role in the development of several types of gliomas. One study demonstrated that SOX7 was downregulated in human glioma, allowing cancer development through upregulated Wnt/  $\beta$ -catenin signaling (Zhao et al., 2016), whereas another study demonstrated that overexpression of SOX7 in high-grade glioma (HGG) promoted cancer development by promoting tumor growth *via* vessel abnormalization (Kim et al., 2018). These somewhat conflicting observations demonstrate that, due to its heavy involvement in regulating several intricately linked developmental and homeostatic functions, SOX7 expression must be delicately balanced. Interestingly, it has also been demonstrated that there is extensive overlap of genetic risk between autism and cancer (Crawley et al., 2016; Crespi, 2011; Gabrielli et al., 2019; Tabarés-Seisdedos & Rubenstein, 2009). SOX7 expression and its interactions may provide additional support for this conjecture, particularly due to its role in vasculature development and Wnt signaling regulation.

## Limitations

The methods performed in this study are not without limitations. Gene expression is a very dynamic process that is not only tissue dependent but also cell type specific and varies depending on the developmental stage and even external factors (Fitzgerald et al., 2004; Hsieh et al., 2000; Lawlor et al., 2017; Shen-Orr et al., 2010; Weyer & Schilling, 2003; Xu et al., 2014). Certainly, these factors affecting genetic expression mean that any autism-related genes that are differentially expressed at different developmental stages or in other varying contexts may be missed. Additionally, differential expression analysis was performed on bulk RNA, whereas it is possible that altered gene expression between autism cases and controls is cell-type specific; knowing the specifics of the expression state of specific cell types that make up key areas of the brain has a better chance of revealing mechanisms behind autism pathogenesis as well as possibly elucidating the pathophysiology behind the vast variety

of ASD subtypes. Gene-based analysis also has some limitations, the most important being the reliance on a reference population for estimating linkage disequilibrium between variants. The similarity of this reference population to the population of study is crucial to the accuracy of many gene-based analyses, including those performed here. As a result, the extent of our findings is limited to European populations, as this was our reference of choice. Future work includes a tighter integration of DNA and RNA information as well as extensions to non-European populations that have been under-researched.

These limitations notwithstanding, the study has considerable strengths. The AT method used in the gene-based GWAS can not only integrate the favorable properties of sum and squared sum tests but also consider LD information among genetic variants. The heatmap of the correlation between genetic variants in SOX7 (Supplementary Figure 1) indicates that rs7005905 and rs7836366, rs10100209 and rs7836366, and rs10100209 and rs7005905 have strong positive linkage disequilibrium (LD) ( $\rho > 0.5$ ); rs4841432 has negative LD with other variants except for rs7009920. The strong LD in SOX7 and the powerful AT method warrant our identification of the autism-associated gene SOX7. The successful replications of SOX7 in the replication data, gene expression data, and the associated biological plausibility underscores the robustness of the finding of the connection between SOX7 and autism. This finding may significantly advance our understanding of the etiology of autism, open new opportunities to reinvigorate stalling autism drug development and increase the accuracy of risk prediction of autism, which makes early autism intervention and prevention possible.

## Conclusion

These findings suggest that SOX7 and its related SOX family genes encode transcription factors that are critical to the downregulation of the canonical Wnt/  $\beta$ -catenin signaling pathway, an important developmental signaling pathway, providing credence to the biologic plausibility of the association between gene SOX7 and autism spectrum disorder.

## Abbreviations

GWAS: genome-wide association studies

ASD: autism spectrum disorder

PGC: Psychiatric Genomics Consortium (PGC)

LD: linkage disequilibrium

ST: Sum test

S2T: Squared sum test

AT: Adaptive test

SRA: sequence read archive

GEO: Gene Expression Omnibus

ATP: Autism Tissue Project

SNPs: single nucleotide polymorphisms

## **Declarations**

## **Ethics approval and consent to participate**

Ethical approval from relevant research ethics committees was granted for each of the original GWAS and RNA-seq data studies and details can be found in the respective publications. Consent to participate: Not applicable.

## **Consent for publication**

Not applicable

## **Availability of data and materials**

The anonymized dataset and materials are available on reasonable request from Moo-yeal Lee (Moo-Yeal.Lee@unt.edu).

## **Competing interests**

The authors declared no potential competing interests with respect to the research, authorship and/or publication of this article.

## **Funding**

This study was financially supported by the National Institutes of Health (NCATS R44TR003491 and NIDDK UH3DK119982) and the University of North Texas (Startup).

## **Authors' contributions**

ML designed and supervised the study; SG and JZ implemented the tests, performed data analyses for all the study; SG and JZ wrote the draft of the manuscript and drew Figures, all authors (SG, JZ, NC, PA, SJ,

and ML) contributed to interpret the results, edit the text and figures. All authors (SG, JZ, NC, PA, SJ, and ML) read and approved the final manuscript.

## Acknowledgements

We would like to thank patients and families from South Florida Autism Center and the University of North Texas Kristin Farmer Autism Center for kindly providing their suggestions and feedback.

## Authors' information

Samantha Gonzales, AHC4 room 228, Department of Biostatistics, Florida International University, Miami, FL 33199 (11200 S.W. 8th Street, AHC4 room 228. Miami, FL 33199), Email: samagonz@fiu.edu

Jane Zizhen Zhao, Miami Dade College Kendall Campus and School for Advanced Studies, Miami, FL 33176 (11011 SW 104 Street, room 4103, Miami, FL 33176) Phone: (305) 237 0510, Email: jane.zhao001@mymdc.net

Na Young Choi, PhD, Postdoc, Department of Biomedical Engineering, University of North Texas, Denton, TX 76207 (Discovery Park K270, 3940 North Elm Street, Denton, TX 76207-7102) Email: nayoung.choi@unt.edu

Prabha Acharya, Department of Biomedical Engineering, University of North Texas, Denton, TX 76207 (Discovery Park K270, 3940 North Elm Street, Denton, TX 76207-7102) Email: PrabhaAcharya@my.unt.edu

Sehoon Jeong, PhD, Assistant Professor, Department of Healthcare Information Technology, Inje University, Gimhae, South Korea, 50834

Dept. Healthcare Information Technology

Inje University

Office: F-903 room, 197 Inje-Ro, Gimhae, South Korea, 50834

Phone: +82-55-320-4131 | E-mail: jeongsh@inje.ac.kr

Homepage: [jeong-lab.wixsite.com/bisnel](http://jeong-lab.wixsite.com/bisnel)

Dr. Moo-Yeal Lee, Ph.D., Associate Professor, Department of Biomedical Engineering, University of North Texas

Office: Discovery Park K240K, 3940 North Elm Street, Denton, TX 76207-7102

Phone: (940) 369-8937

## References

1. Abu-Elneel, K., Liu, T., Gazzaniga, F. S., Nishimura, Y., Wall, D. P., Geschwind, D. H., Lao, K., & Kosik, K. S. (2008). Heterogeneous dysregulation of microRNAs across the autism spectrum. *neurogenetics*, 9(3), 153-161. <https://doi.org/10.1007/s10048-008-0133-5>
2. Alonso-Gonzalez, A., Calaza, M., Rodriguez-Fontenla, C., & Carracedo, A. (2019). Novel Gene-Based Analysis of ASD GWAS: Insight Into the Biological Role of Associated Genes. *Frontiers in Genetics*, 10. <https://www.frontiersin.org/articles/10.3389/fgene.2019.00733>
3. Andrews, S. (2010). *FastQC: a quality control tool for high throughput sequence data*. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.
4. Aspra, Q., Cabrera-Mendoza, B., Morales-Marín, M. E., Márquez, C., Chicalote, C., Ballesteros, A., Aguilar, M., Castro, X., Gómez-Cotero, A., Balboa-Verduzco, A. M., Albores-Gallo, L., Nafate-López, O., Marcín-Salazar, C. A., Sánchez, P., Lanzagorta-Piñol, N., López-Armenta, F. O., & Nicolini, H. (2022). Epigenome-Wide Analysis Reveals DNA Methylation Alteration in ZFP57 and Its Target RASGFR2 in a Mexican Population Cohort with Autism. *Children*, 9(4), 462. <https://doi.org/10.3390/children9040462>
5. Baranova, A., Wang, J., Cao, H., Chen, J.-H., Chen, J., Chen, M., Ni, S., Xu, X., Ke, X., Xie, S., Sun, J., & Zhang, F. (2022). Shared genetics between autism spectrum disorder and attention-deficit/hyperactivity disorder and their association with extraversion. *Psychiatry Research*, 314, 114679. <https://doi.org/10.1016/j.psychres.2022.114679>
6. Berto, S., Treacher, A. H., Caglayan, E., Luo, D., Haney, J. R., Gandal, M. J., Geschwind, D. H., Montillo, A. A., & Konopka, G. (2022). Association between resting-state functional brain connectivity and gene expression is altered in autism spectrum disorder. *Nature Communications*, 13(1), 3328. <https://doi.org/10.1038/s41467-022-31053-5>
7. Bjørklund, G., Meguid, N. A., El-Bana, M. A., Tinkov, A. A., Saad, K., Dadar, M., Hemimi, M., Skalny, A. V., Hosnedlová, B., Kizek, R., Osredkar, J., Urbina, M. A., Fabjan, T., El-Houfey, A. A., Kałużna-Czaplińska, J., Gątarek, P., & Chirumbolo, S. (2020). Oxidative Stress in Autism Spectrum Disorder. *Molecular Neurobiology*, 57(5), 2314-2332. <https://doi.org/10.1007/s12035-019-01742-2>
8. Bourgeron, T. (2015). From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nature Reviews Neuroscience*, 16(9), 551-563.
9. Brandenburg, C., Soghomonian, J.-J., Zhang, K., Sulkaj, I., Randolph, B., Kachadoorian, M., & Blatt, G. J. (2020). Increased Dopamine Type 2 Gene Expression in the Dorsal Striatum in Individuals With Autism Spectrum Disorder Suggests Alterations in Indirect Pathway Signaling and Circuitry. *Frontiers in Cellular Neuroscience*, 14. <https://www.frontiersin.org/articles/10.3389/fncel.2020.577858>
10. Brannan, K., Kim, H., Erickson, B., Glover-Cutter, K., Kim, S., Fong, N., Kiemele, L., Hansen, K., Davis, R., Lykke-Andersen, J., & Bentley, D. L. (2012). mRNA Decapping Factors and the Exonuclease Xrn2



- Function in Widespread Premature Termination of RNA Polymerase II Transcription. *Molecular Cell*, 46(3), 311-324. <https://doi.org/10.1016/j.molcel.2012.03.006>
11. Bryan N. Howie, P. D., Jonathan Marchini. (2009). A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies | PLOS Genetics. <https://doi.org/10.1371/journal.pgen.1000529>
  12. Buitrago, D., Labrador, M., Arcon, J. P., Lema, R., Flores, O., Esteve-Codina, A., Blanc, J., Villegas, N., Bellido, D., Gut, M., Dans, P. D., Heath, S. C., Gut, I. G., Brun Heath, I., & Orozco, M. (2021). Impact of DNA methylation on 3D genome structure. *Nature Communications*, 12(1), 3243. <https://doi.org/10.1038/s41467-021-23142-8>
  13. Buxbaum, J. D. (2022). Multiple rare variants in the etiology of autism spectrum disorders. *Dialogues in clinical neuroscience*.
  14. Caracci, M. O., Avila, M. E., Espinoza-Cavieres, F. A., López, H. R., Ugarte, G. D., & De Ferrari, G. V. (2021). Wnt/ $\beta$ -Catenin-Dependent Transcription in Autism Spectrum Disorders. *Frontiers in Molecular Neuroscience*, 14. <https://www.frontiersin.org/articles/10.3389/fnmol.2021.764756>
  15. Casanova, M. F. (2007). The Neuropathology of Autism. *Brain Pathology*, 17(4), 422-433. <https://doi.org/10.1111/j.1750-3639.2007.00100.x>
  16. Chang, C. C., Complete Genomics, S. C., 94043 Mountain View, CA, USA, BGI Cognitive Genomics Lab, B. N., Bei Shan Industrial Zone, Yantian District, 518083 Shenzhen, China, Chow, C. C., Mathematical Biology Section, N. L., National Institutes of Health, 20892 Bethesda, MD, USA, Tellier, L. C., BGI Cognitive Genomics Lab, B. N., Bei Shan Industrial Zone, Yantian District, 518083 Shenzhen, China, Bioinformatics Centre, U. o. C., 2200 Copenhagen, Denmark, Vattikuti, S., Mathematical Biology Section, N. L., National Institutes of Health, 20892 Bethesda, MD, USA, Purcell, S. M., Stanley Center for Psychiatric Research, B. I. o. M. a. H., 02142 Cambridge, MA, USA, Division of Psychiatric Genomics, D. o. P., Icahn School of Medicine at Mount Sinai, 10029 New York, NY, USA, Institute for Genomics and Multiscale Biology, I. S. o. M. a. M. S., 10029 New York, NY, USA, Analytic and Translational Genetics Unit, P. a. N. G. U., Massachusetts General Hospital, 02114 Boston, MA, USA, Lee, J. J., Mathematical Biology Section, N. L., National Institutes of Health, 20892 Bethesda, MD, USA, & Department of Psychology, U. o. M. T. C., 55455 Minneapolis, MN, USA. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4(1). <https://doi.org/10.1186/s13742-015-0047-8>
  17. Chauhan, A., & Chauhan, V. (2006). Oxidative stress in autism. *Pathophysiology*, 13(3), 171-181. <https://doi.org/10.1016/j.pathophys.2006.05.007> (Oxidative Stress and Metabolic diseases)
  18. Chien, W.-H., Gau, S.-F., Chen, C.-H., Tsai, W.-C., Wu, Y.-Y., Chen, P.-H., Shang, C.-Y., & Chen, C.-H. (2013). Increased gene expression of FOXP1 in patients with autism spectrum disorders. *Molecular autism*, 4(1), 23. <https://doi.org/10.1186/2040-2392-4-23>
  19. Consortium, T. A. S. D. W. G. o. t. P. G. (2017). Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24. 32 and a significant overlap with schizophrenia. *Molecular autism*, 8, 1-17.

20. Crawley, J. N., Heyer, W.-D., & LaSalle, J. M. (2016). Autism and Cancer Share Risk Genes, Pathways, and Drug Targets. *Trends in Genetics*, 32(3), 139-146. <https://doi.org/10.1016/j.tig.2016.01.001>
21. Crespi, B. (2011). Autism and cancer risk. *Autism Research*, 4(4), 302-310. <https://doi.org/10.1002/aur.208>
22. de la Torre-Ubieta, L., Won, H., Stein, J. L., & Geschwind, D. H. (2016). Advancing the understanding of autism disease mechanisms through genetics. *Nature Medicine*, 22(4), 345-361. <https://doi.org/10.1038/nm.4071>
23. Delaneau, O., Marchini, J., & Zagury, J.-F. (2011). A linear complexity phasing method for thousands of genomes [OriginalPaper]. *Nature Methods*, 9(2), 179-181. <https://doi.org/doi:10.1038/nmeth.1785>
24. Devlin, B., Kelsoe, J. R., Sklar, P., Daly, M. J., O'Donovan, M. C., Craddock, N., Sullivan, P. F., Smoller, J. W., Kendler, K. S., & others. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature genetics*, 45(9), 984-994.
25. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15-21. <https://doi.org/10.1093/bioinformatics/bts635>
26. Eaton, J. D., & West, S. (2018). An end in sight? Xrn2 and transcriptional termination by RNA polymerase II. *Transcription*, 9(5), 321-326. <https://doi.org/10.1080/21541264.2018.1498708>
27. El Khouri, E., Ghomid, J., Haye, D., Giuliano, F., Drevillon, L., Briand-Suleau, A., De La Grange, P., Nau, V., Gaillon, T., Bienvenu, T., Jacquemin-Sablon, H., Goossens, M., Amselem, S., & Giurgea, I. (2021). Wnt/ $\beta$ -catenin pathway and cell adhesion deregulation in CSDE1-related intellectual disability and autism spectrum disorders. *Molecular psychiatry*, 26(7), 3572-3585. <https://doi.org/10.1038/s41380-021-01072-7>
28. Emanuele, E., Orsi, P., Barale, F., di Nemi, S. U., Bertona, M., & Politi, P. (2010). Serum levels of vascular endothelial growth factor and its receptors in patients with severe autism. *Clinical Biochemistry*, 43(3), 317-319. <https://doi.org/10.1016/j.clinbiochem.2009.10.005>
29. Ewels, P., Magnusson, M., Lundin, S., & K  ller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047-3048. <https://doi.org/10.1093/bioinformatics/btw354>
30. FASTX-Toolkit. [http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)
31. Ferland, R. J., Cherry, T. J., Preware, P. O., Morrissey, E. E., & Walsh, C. A. (2003). Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. *Journal of Comparative Neurology*, 460(2), 266-279. <https://doi.org/10.1002/cne.10654>
32. Fitzgerald, J. B., Jin, M., Dean, D., Wood, D. J., Zheng, M. H., & Grodzinsky, A. J. (2004). Mechanical Compression of Cartilage Explants Induces Multiple Time-dependent Gene Expression Patterns and Involves Intracellular Calcium and Cyclic AMP \*. *Journal of Biological Chemistry*, 279(19), 19502-19511. <https://doi.org/10.1074/jbc.M400437200>
33. Fombonne, E. (2009). Epidemiology of pervasive developmental disorders. *Pediatric research*, 65(6), 591-598.

34. Francois, M., Koopman, P., & Beltrame, M. (2010). SoxF genes: Key players in the development of the cardio-vascular system. *The International Journal of Biochemistry & Cell Biology*, 42(3), 445-448. <https://doi.org/10.1016/j.biocel.2009.08.017> (SOX Transcription Factors)
35. Frankish, A., Diekhans, M., Ferreira, A., Johnson, R., Jungreis, I., Loveland, J., Mudge, J., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Carbonell, S., Chrast, J., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I., García, G. C., Gonzalez, J., Grego, T., Hardy, M., Hourlier, T., Hunt, T., Izuogu, O., Lagarde, J., Martin, F., Martínez, L., Mohanan, S., Muir, P., Navarro, F., Parker, A., Pei, B., Pozo, F., Ruffier, M., Schmitt, B., Stapleton, E., Suner, M., Sycheva, I., Uszczynska-Ratajczak, B., Xu, J., Yates, A., Zerbino, D., Zhang, Y., Aken, B., Choudhary, J., Gerstein, M., Guigó, R., Hubbard, T., Kellis, M., Paten, B., Reymond, A., Tress, M., & Flicek, P. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic acids research*, 47(D1). <https://doi.org/10.1093/nar/gky955>
36. Gabrielli, A. P., Manzardo, A. M., & Butler, M. G. (2019). GeneAnalytics Pathways and Profiling of Shared Autism and Cancer Genes. *International Journal of Molecular Sciences*, 20(5), 1166. <https://doi.org/10.3390/ijms20051166>
37. Gaugler, T., Klei, L., Sanders, S. J., Bodea, C. A., Goldberg, A. P., Lee, A. B., Mahajan, M., Manaa, D., Pawitan, Y., Reichert, J., & others. (2014). Most genetic risk for autism resides with common variation. *Nature genetics*, 46(8), 881-885.
38. Ghahramani Seno, M. M., Hu, P., Gwady, F. G., Pinto, D., Marshall, C. R., Casallo, G., & Scherer, S. W. (2011). Gene and miRNA expression profiles in autism spectrum disorders. *Brain Research*, 1380, 85-97. <https://doi.org/10.1016/j.brainres.2010.09.046> (The Emerging Neuroscience of Autism Spectrum Disorders)
39. Gillberg, C., Fernell, E., & Minnis, H. (2014). Early symptomatic syndromes eliciting neurodevelopmental clinical examinations. In: Hindawi.
40. Gozal, E., Jagadapillai, R., Cai, J., & Barnes, G. N. (2021). Potential crosstalk between sonic hedgehog-WNT signaling and neurovascular molecules: Implications for blood–brain barrier integrity in autism spectrum disorder. *Journal of Neurochemistry*, 159(1), 15-28. <https://doi.org/10.1111/jnc.15460>
41. Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R. K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O. A., Anney, R., & others. (2019). Identification of common genetic risk variants for autism spectrum disorder. *Nature genetics*, 51(3), 431-444.
42. Guo, B., & Wu, B. (2018). Statistical methods to detect novel genetic variants using publicly available GWAS summary data. *Computational Biology and Chemistry*, 74, 76-79. <https://doi.org/https://doi.org/10.1016/j.compbiolchem.2018.02.016>
43. Hannon, E., Schendel, D., Ladd-Acosta, C., Grove, J., Agerbo, E., Als, T. D., Belliveau, R., Bybjerg-Grauholm, J., Bækved-Hansen, M., Børglum, A., Cerrato, F., Christensen, J., Chambert, K., Churchhouse, C., Daly, M., Demontis, D., Dumont, A., Goldstein, J., Hansen, C., Hauberg, M., Hougaard, D., Howrigan, D., Huang, H., Maller, J., Martin, A., Martin, J., Mattheisen, M., Moran, J., Mors, O., Mortensen, P., Neale, B., Nordentoft, M., Nyegaard, M., Pallesen, J., Palmer, D., Pedersen, C., Pedersen,

- M., Poterba, T., Poulsen, J., Qvist, P., Ripke, S., Robinson, E., Satterstrom, K., Stevens, C., Turley, P., Walters, R., Werge, T., Hansen, C. S., Andrews, S. V., Hougaard, D. M., Bresnahan, M., Hollegaard, M. V., Bækvad-Hansen, M., Hornig, M., Mortensen, P. B., Børglum, A. D., Pedersen, M. G., Buxbaum, J., Daniele Fallin, M., Reichenberg, A., Mill, J., & i, P.-B. A. S. D. G. (2018). Elevated polygenic burden for autism is associated with differential DNA methylation at birth. *Genome Medicine*, 10(1), 19. <https://doi.org/10.1186/s13073-018-0527-4>
44. Hicks, S. D., & Middleton, F. A. (2016). A Comparative Review of microRNA Expression Patterns in Autism Spectrum Disorder. *Frontiers in Psychiatry*, 7. <https://www.frontiersin.org/articles/10.3389/fpsy.2016.00176>
  45. Hormozdiari, F., Penn, O., Borenstein, E., & Eichler, E. E. (2015). The discovery of integrated gene networks for autism and related disorders. *Genome Research*, 25(1), 142-154. <https://doi.org/10.1101/gr.178855.114>
  46. Hsieh, A. H., Tsai, C. M. H., Ma, Q.-J., Lin, T., Banes, A. J., Villarreal, F. J., Akeson, W. H., & Paul Sung, K. L. (2000). Time-dependent increases in type-III collagen gene expression in medial collateral ligament fibroblasts under cyclic strains. *Journal of Orthopaedic Research*, 18(2), 220-227. <https://doi.org/10.1002/jor.1100180209>
  47. Huang, K., Wu, Y., Shin, J., Zheng, Y., Siahpirani, A. F., Lin, Y., Ni, Z., Chen, J., You, J., Keles, S., Wang, D., Roy, S., & Lu, Q. (2021). Transcriptome-wide transmission disequilibrium analysis identifies novel risk genes for autism spectrum disorder. *PLOS Genetics*, 17(2), e1009309. <https://doi.org/10.1371/journal.pgen.1009309>
  48. Jamain, S., Betancur, C., Quach, H., Philippe, A., Fellous, M., Giros, B., Gillberg, C., Leboyer, M., & Bourgeron, T. (2002). Linkage and association of the glutamate receptor 6 gene with autism. *Molecular psychiatry*, 7(3), 302-310.
  49. Jamain, S., Quach, H., Betancur, C., Råstam, M., Colineaux, C., Gillberg, I. C., Soderstrom, H., Giros, B., Leboyer, M., Gillberg, C., & others. (2003). Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nature genetics*, 34(1), 27-29.
  50. Johnson, F. B. (2011). PinX1 the tail on the chromosome. *The Journal of Clinical Investigation*, 121(4), 1242-1244. <https://doi.org/10.1172/JCI57024>
  51. Katoh, M. (2002). Expression of human SOX7 in normal tissues and tumors. *International Journal of Molecular Medicine*, 9(4), 363-368. <https://doi.org/10.3892/ijmm.9.4.363>
  52. Keshet, I., Lieman-Hurwitz, J., & Cedar, H. (1986). DNA methylation affects the formation of active chromatin. *Cell*, 44(4), 535-543. [https://doi.org/10.1016/0092-8674\(86\)90263-1](https://doi.org/10.1016/0092-8674(86)90263-1)
  53. Kim, I.-K., Kim, K., Lee, E., Oh, D. S., Park, C. S., Park, S., Yang, J. M., Kim, J.-H., Kim, H.-S., Shima, D. T., Kim, J. H., Hong, S. H., Cho, Y. H., Kim, Y. H., Park, J. B., Koh, G. Y., Ju, Y. S., Lee, H. K., Lee, S., & Kim, I. (2018). Sox7 promotes high-grade glioma by increasing VEGFR2-mediated vascular abnormality. *The Journal of Experimental Medicine*, 215(3), 963-983. <https://doi.org/10.1084/jem.20170123>
  54. Kim, K., Kim, I.-K., Yang, J. M., Lee, E., Koh, B. I., Song, S., Park, J., Lee, S., Choi, C., Kim, J. W., Kubota, Y., Koh, G. Y., & Kim, I. (2016). SoxF Transcription Factors Are Positive Feedback Regulators of VEGF

- Signaling. *Circulation Research*, 119(7), 839-852. <https://doi.org/10.1161/CIRCRESAHA.116.308483>
55. Kinjo, E. R., Higa, G. S. V., de Sousa, E., Casado, O. A. N., Damico, M. V., Britto, L. R. G., & Kihara, A. H. (2013). A possible new mechanism for the control of miRNA expression in neurons. *Experimental Neurology*, 248, 546-558. <https://doi.org/10.1016/j.expneurol.2013.07.022>
  56. Klomp, J., Hyun, J., Klomp, J. E., Pajcini, K., Rehman, J., & Malik, A. B. (2020). Comprehensive transcriptomic profiling reveals SOX7 as an early regulator of angiogenesis in hypoxic human endothelial cells. *Journal of Biological Chemistry*, 295(15), 4796-4808. <https://doi.org/10.1074/jbc.RA119.011822>
  57. Kwan, V., Unda, B. K., & Singh, K. K. (2016). Wnt signaling networks in autism spectrum disorder and intellectual disability. *Journal of Neurodevelopmental Disorders*, 8(1), 1-10. <https://doi.org/10.1186/s11689-016-9176-3>
  58. Lawlor, N., George, J., Bolisetty, M., Kursawe, R., Sun, L., Sivakamasundari, V., Kycia, I., Robson, P., & Stitzel, M. L. (2017). Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. *Genome Research*, 27(2), 208-222. <https://doi.org/10.1101/gr.212720.116>
  59. Leblond, C. S., Cliquet, F., Carton, C., Huguet, G., Mathieu, A., Kergrohen, T., Buratti, J., Lemièrre, N., Cuisset, L., Bienvenu, T., & others. (2019). Both rare and common genetic variants contribute to autism in the Faroe Islands. *NPJ genomic medicine*, 4(1), 1-10.
  60. Levy, D., Ronemus, M., Yamrom, B., Lee, Y.-h., Leotta, A., Kendall, J., Marks, S., Lakshmi, B., Pai, D., Ye, K., & others. (2011). Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron*, 70(5), 886-897.
  61. Lewis, J., & Bird, A. (1991). DNA methylation and chromatin structure. *FEBS Letters*, 285(2), 155-159. [https://doi.org/10.1016/0014-5793\(91\)80795-5](https://doi.org/10.1016/0014-5793(91)80795-5)
  62. Li, X., Zou, H., & Brown, W. T. (2012). Genes associated with autism spectrum disorder. *Brain research bulletin*, 88(6), 543-552.
  63. Lilly, A. J., Mazan, A., Scott, D. A., Lacaud, G., & Kouskoff, V. (2017). SOX7 expression is critically required in FLK1-expressing cells for vasculogenesis and angiogenesis during mouse embryonic development. *Mechanisms of Development*, 146, 31-41. <https://doi.org/10.1016/j.mod.2017.05.004>
  64. Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550. <https://doi.org/10.1186/s13059-014-0550-8>
  65. MacDonald, B. T., Tamai, K., & He, X. (2009). Wnt/ $\beta$ -Catenin Signaling: Components, Mechanisms, and Diseases. *Developmental Cell*, 17(1), 9-26. <https://doi.org/10.1016/j.devcel.2009.06.016>
  66. Marques, A. R., Santos, J. X., Martiniano, H., Vilela, J., Rasga, C., Romão, L., & Vicente, A. M. (2022). Gene Variants Involved in Nonsense-Mediated mRNA Decay Suggest a Role in Autism Spectrum Disorder. *Biomedicines*, 10(3), 665. <https://doi.org/10.3390/biomedicines10030665>
  67. Melke, J., Goubran Botros, H., Chaste, P., Betancur, C., Nygren, G., Anckarsäter, H., Rastam, M., Ståhlberg, O., Gillberg, I. C., Delorme, R., & others. (2008). Abnormal melatonin synthesis in autism

- spectrum disorders. *Molecular psychiatry*, 13(1), 90-98.
68. Nagarajan, V. K., Jones, C. I., Newbury, S. F., & Green, P. J. (2013). XRN 5'→3' exoribonucleases: Structure, mechanisms and functions. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1829(6), 590-603. <https://doi.org/10.1016/j.bbagr.2013.03.005> (RNA Decay Mechanisms)
  69. Nussbacher, J. K., Tabet, R., Yeo, G. W., & Lagier-Tourenne, C. (2019). Disruption of RNA Metabolism in Neurological Diseases and Emerging Therapeutic Interventions. *Neuron*, 102(2), 294-320. <https://doi.org/10.1016/j.neuron.2019.03.014>
  70. O'Roak, B. J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B. P., Levy, R., Ko, A., Lee, C., Smith, J. D., & others. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*, 485(7397), 246-250.
  71. Oshimori, N., Ohsugi, M., & Yamamoto, T. (2006). The Plk1 target Kizuna stabilizes mitotic centrosomes to ensure spindle bipolarity. *Nature Cell Biology*, 8(10), 1095-1101. <https://doi.org/10.1038/ncb1474>
  72. Ouellette, J., Toussay, X., Comin, C. H., Costa, L. d. F., Ho, M., Lacalle-Aurioles, M., Freitas-Andrade, M., Liu, Q. Y., Leclerc, S., Pan, Y., Liu, Z., Thibodeau, J.-F., Yin, M., Carrier, M., Morse, C. J., Dyken, P. V., Bergin, C. J., Baillet, S., Kennedy, C. R., Tremblay, M.-È., Benoit, Y. D., Stanford, W. L., Burger, D., Stewart, D. J., & Lacoste, B. (2020). Vascular contributions to 16p11.2 deletion autism syndrome modeled in mice. *Nature Neuroscience*, 23(9), 1090-1101. <https://doi.org/10.1038/s41593-020-0663-1>
  73. Packer, A. (2016). Neocortical neurogenesis and the etiology of autism spectrum disorder. *Neuroscience & Biobehavioral Reviews*, 64, 185-195. <https://doi.org/10.1016/j.neubiorev.2016.03.002>
  74. Páez, M. T., Yamamoto, T., Hayashi, K.-i., Yasuda, T., Harada, N., Matsumoto, N., Kurosawa, K., Furutani, Y., Asakawa, S., Shimizu, N., & Matsuoka, R. (2008). Two patients with atypical interstitial deletions of 8p23.1: Mapping of phenotypical traits. *American Journal of Medical Genetics Part A*, 146A(9), 1158-1165. <https://doi.org/10.1002/ajmg.a.32205>
  75. Pain, O., Pocklington, A. J., Holmans, P. A., Bray, N. J., O'Brien, H. E., Hall, L. S., Pardiñas, A. F., O'Donovan, M. C., Owen, M. J., & Anney, R. (2019). Novel Insight Into the Etiology of Autism Spectrum Disorder Gained by Integrating Expression Data With Genome-wide Association Statistics. *Biological Psychiatry*, 86(4), 265-273. <https://doi.org/10.1016/j.biopsych.2019.04.034> (Autism Spectrum Disorder: Mechanisms and Features)
  76. Parikshak, N. N., Swarup, V., Belgard, T. G., Irimia, M., Ramaswami, G., Gandal, M. J., Hartl, C., Leppa, V., Ubieta, L. d. I. T., Huang, J., Lowe, J. K., Blencowe, B. J., Horvath, S., & Geschwind, D. H. (2016). Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature*, 540(7633), 423-427. <https://doi.org/10.1038/nature20612>
  77. Peyre, H., Schoeler, T., Liu, C., Williams, C. M., Hoertel, N., Havdahl, A., & Pingault, J.-B. (2020). Combining multivariate genomic approaches to elucidate the comorbidity between ASD and ADHD.

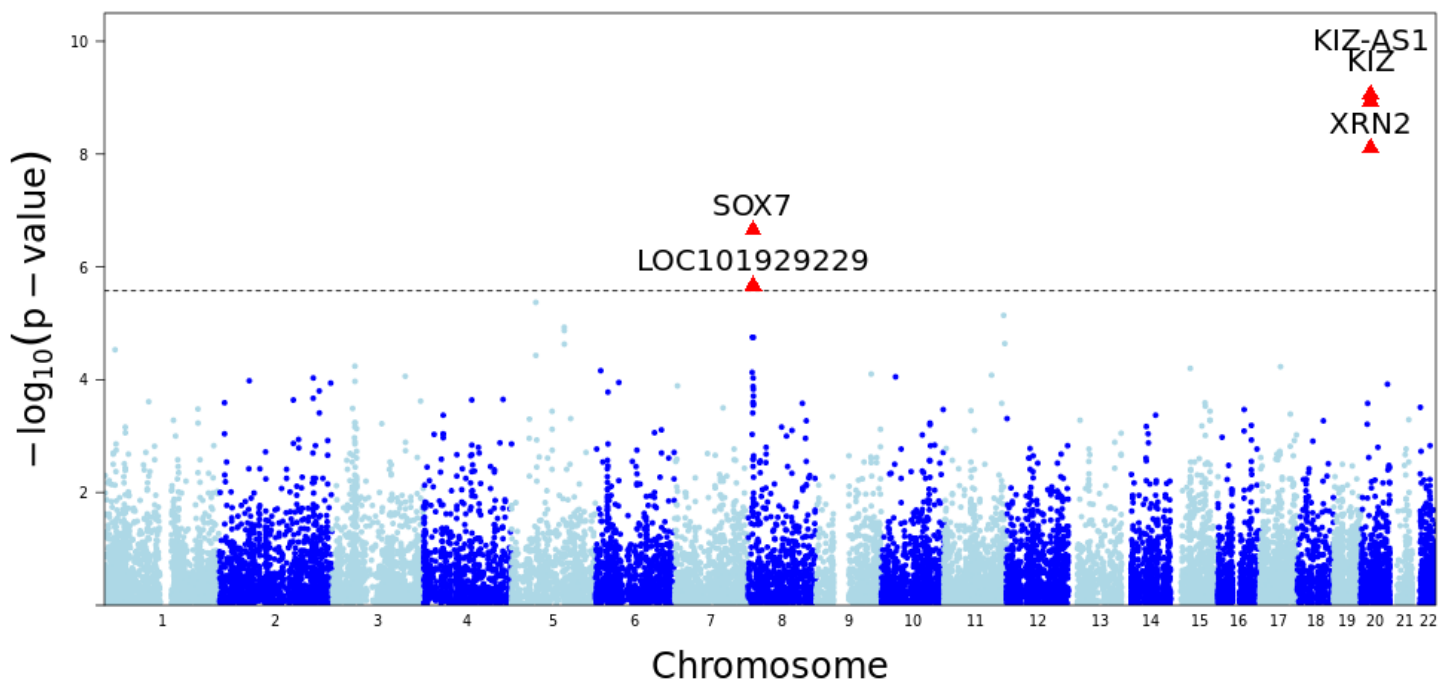
78. Pramparo, T., Lombardo, M. V., Campbell, K., Barnes, C. C., Marinero, S., Solso, S., Young, J., Mayo, M., Dale, A., Ahrens-Barbeau, C., Murray, S. S., Lopez, L., Lewis, N., Pierce, K., & Courchesne, E. (2015). Cell cycle networks link gene expression dysregulation, mutation, and brain maldevelopment in autistic toddlers. *Molecular Systems Biology*, 11(12), 841. <https://doi.org/10.15252/msb.20156108>
79. Putri, G. H., Anders, S., Pyl, P. T., Pimanda, J. E., & Zanini, F. (2022). Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics*, 38(10), 2943-2945. <https://doi.org/10.1093/bioinformatics/btac166>
80. Quesnel-Vallières, M., Weatheritt, R. J., Cordes, S. P., & Blencowe, B. J. (2019). Autism spectrum disorder: insights into convergent mechanisms from transcriptomics. *Nature Reviews Genetics*, 20(1), 51-63. <https://doi.org/10.1038/s41576-018-0066-2>
81. Rahman, M. R., Petralia, M. C., Ciurleo, R., Bramanti, A., Fagone, P., Shahjaman, M., Wu, L., Sun, Y., Turanli, B., Arga, K. Y., Islam, M. R., Islam, T., & Nicoletti, F. (2020). Comprehensive Analysis of RNA-Seq Gene Expression Profiling of Brain Transcriptomes Reveals Novel Genes, Regulators, and Pathways in Autism Spectrum Disorder. *Brain Sciences*, 10(10), 747. <https://doi.org/10.3390/brainsci10100747>
82. Reis, M., & Liebner, S. (2013). Wnt signaling in the vasculature. *Experimental Cell Research*, 319(9), 1317-1323. <https://doi.org/10.1016/j.yexcr.2012.12.023> (Special Issue: Endothelial Biology)
83. Robinson, E. B., St Pourcain, B., Anttila, V., Kosmicki, J. A., Bulik-Sullivan, B., Grove, J., Maller, J., Samocha, K. E., Sanders, S. J., Ripke, S., & others. (2016). Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nature genetics*, 48(5), 552-555.
84. Roohi, J., Montagna, C., Tegay, D. H., Palmer, L. E., DeVincent, C., Pomeroy, J. C., Christian, S. L., Nowak, N., & Hatchwell, E. (2009). Disruption of contactin 4 in three subjects with autism spectrum disorder. *Journal of medical genetics*, 46(3), 176-182.
85. Sanders, S. J., Ercan-Sencicek, A. G., Hus, V., Luo, R., Murtha, M. T., Moreno-De-Luca, D., Chu, S. H., Moreau, M. P., Gupta, A. R., Thomson, S. A., & others. (2011). Multiple recurrent de novo CNVs, including duplications of the 7q11. 23 Williams syndrome region, are strongly associated with autism. *Neuron*, 70(5), 863-885.
86. Sanders, S. J., He, X., Willsey, A. J., Ercan-Sencicek, A. G., Samocha, K. E., Cicek, A. E., Murtha, M. T., Bal, V. H., Bishop, S. L., Dong, S., & others. (2015). Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron*, 87(6), 1215-1233.
87. Sanders, S. J., Murtha, M. T., Gupta, A. R., Murdoch, J. D., Raubeson, M. J., Willsey, A. J., Ercan-Sencicek, A. G., DiLullo, N. M., Parikshak, N. N., Stein, J. L., & others. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*, 485(7397), 237-241.
88. Sandin, S., Lichtenstein, P., Kuja-Halkola, R., Hultman, C., Larsson, H., & Reichenberg, A. (2017). The heritability of autism spectrum disorder. *Jama*, 318(12), 1182-1184.

89. Satterstrom, F. K., Kosmicki, J. A., Wang, J., Breen, M. S., De Rubeis, S., An, J.-Y., Peng, M., Collins, R., Grove, J., Klei, L., & others. (2020). Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell*, 180(3), 568-584.
90. Shen-Orr, S. S., Tibshirani, R., Khatri, P., Bodian, D. L., Staedtler, F., Perry, N. M., Hastie, T., Sarwal, M. M., Davis, M. M., & Butte, A. J. (2010). Cell type-specific gene expression differences in complex tissues. *Nature Methods*, 7(4), 287-289. <https://doi.org/10.1038/nmeth.1439>
91. SRA Toolkit. *GitHub*. <https://github.com/ncbi/sra-tools/wiki/01.-Downloading-SRA-Toolkit>  
<https://github.com/ncbi/sra-tools>
92. Stenman, J. M., Rajagopal, J., Carroll, T. J., Ishibashi, M., McMahon, J., & McMahon, A. P. (2008). Canonical Wnt Signaling Regulates Organ-Specific Assembly and Differentiation of CNS Vasculature. *Science*, 322(5905), 1247-1250. <https://doi.org/10.1126/science.1164594>
93. Tabarés-Seisdedos, R., & Rubenstein, J. L. R. (2009). Chromosome 8p as a potential hub for developmental neuropsychiatric disorders: implications for schizophrenia, autism and cancer. *Molecular psychiatry*, 14(6), 563-589. <https://doi.org/10.1038/mp.2009.2>
94. Takash, W., Cañizares, J., Bonneaud, N., Poulat, F., Mattéi, M. G., Jay, P., & Berta, P. (2001). SOX7 transcription factor: sequence, chromosomal localisation, expression, transactivation and interference with Wnt signalling. *Nucleic acids research*, 29(21), 4274-4283. <https://doi.org/10.1093/nar/29.21.4274>
95. Tick, B., Bolton, P., Happé, F., Rutter, M., & Rijdsdijk, F. (2016). Heritability of autism spectrum disorders: a meta-analysis of twin studies. *Journal of Child Psychology and Psychiatry*, 57(5), 585-595.
96. Vallée, A., Vallée, J.-N., & Lecarpentier, Y. (2019). PPAR $\gamma$  agonists: potential treatment for autism spectrum disorder by inhibiting the canonical WNT/ $\beta$ -catenin pathway. *Molecular psychiatry*, 24(5), 643-652. <https://doi.org/10.1038/s41380-018-0131-4>
97. Voineagu, I., Wang, X., Johnston, P., Lowe, J., Tian, Y., Horvath, S., Mill, J., Cantor, R., Blencowe, B., & Geschwind, D. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*, 474(7351). <https://doi.org/10.1038/nature10110>
98. Wang, M., & Pestov, D. G. (2011). 5'-end surveillance by Xrn2 acts as a shared mechanism for mammalian pre-rRNA maturation and decay. *Nucleic acids research*, 39(5), 1811-1822. <https://doi.org/10.1093/nar/gkq1050>
99. Wat, M. J., Shchelochkov, O. A., Holder, A. M., Berman, A. M., Dagli, A., Bacino, C., Scaglia, F., Zori, R. T., Cheung, S. W., Scott, D. A., & Kang, S.-H. L. (2009). Chromosome 8p23.1 deletions as a cause of complex congenital heart defects and diaphragmatic hernia. *American Journal of Medical Genetics Part A*, 149A(8), 1661-1677. <https://doi.org/10.1002/ajmg.a.32896>
100. Weber, A., Köhler, A., Hahn, A., & Müller, U. (2014). 8p23.1 duplication syndrome: narrowing of critical interval to 1.80 Mbp. *Molecular Cytogenetics*, 7(1), 94. <https://doi.org/10.1186/s13039-014-0094-3>
101. West, S., Gromak, N., & Proudfoot, N. J. (2004). Human 5'  $\rightarrow$  3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*, 432(7016), 522-525. <https://doi.org/10.1038/nature03035>



102. Weyer, A., & Schilling, K. (2003). Developmental and cell type-specific expression of the neuronal marker NeuN in the murine cerebellum. *Journal of Neuroscience Research*, 73(3), 400-409. <https://doi.org/10.1002/jnr.10655>
103. Wu, Y. E., Parikshak, N. N., Belgard, T. G., & Geschwind, D. H. (2016). Genome-wide, integrative analysis implicates microRNA dysregulation in autism spectrum disorder. *Nature Neuroscience*, 19(11), 1463-1476. <https://doi.org/10.1038/nn.4373>
104. Xu, X., Wells, A. B., O'Brien, D. R., Nehorai, A., & Dougherty, J. D. (2014). Cell Type-Specific Expression Analysis to Identify Putative Cellular Mechanisms for Neurogenetic Disorders. *Journal of Neuroscience*, 34(4), 1420-1431. <https://doi.org/10.1523/JNEUROSCI.4488-13.2014>
105. Yang, Z., Wu, H., Lee, P. H., Tsetsos, F., Davis, L. K., Yu, D., Lee, S. H., Dalsgaard, S., Haavik, J., Barta, C., Zayats, T., Eapen, V., Wray, N. R., Devlin, B., Daly, M., Neale, B., Børglum, A. D., Crowley, J. J., Scharf, J., Mathews, C. A., Faraone, S. V., Franke, B., Mattheisen, M., Smoller, J. W., & Paschou, P. (2021). Investigating Shared Genetic Basis Across Tourette Syndrome and Comorbid Neurodevelopmental Disorders Along the Impulsivity-Compulsivity Spectrum. *Biological Psychiatry*, 90(5), 317-327. <https://doi.org/10.1016/j.biopsych.2020.12.028> (BPS 90/5 Cortical Development and Autism Spectrum Disorder)
106. Yao, Y., Walsh, W. J., McGinnis, W. R., & Praticò, D. (2006). Altered Vascular Phenotype in Autism: Correlation With Oxidative Stress. *Archives of Neurology*, 63(8), 1161-1164. <https://doi.org/10.1001/archneur.63.8.1161>
107. Yasuda, Y., Hashimoto, R., Yamamori, H., Ohi, K., Fukumoto, M., Umeda-Yano, S., Mohri, I., Ito, A., Taniike, M., & Takeda, M. (2011). Gene expression analysis in lymphoblasts derived from patients with autism spectrum disorder. *Molecular autism*, 2(1), 9. <https://doi.org/10.1186/2040-2392-2-9>
108. Zhang, Y., Sun, Y., Wang, F., Wang, Z., Peng, Y., & Li, R. (2012). Downregulating the Canonical Wnt/ $\beta$ -catenin Signaling Pathway Attenuates the Susceptibility to Autism-like Phenotypes by Decreasing Oxidative Stress. *Neurochemical Research*, 37(7), 1409-1419. <https://doi.org/10.1007/s11064-012-0724-2>
109. Zhao, T., Yang, H., Tian, Y., Xie, Q., Lu, Y., Wang, Y., Su, N., Dong, B., Liu, X., Wang, C., Jiang, C., & Liu, X. (2016). SOX7 is associated with the suppression of human glioma by HMG-box dependent regulation of Wnt/ $\beta$ -catenin signaling. *Cancer Letters*, 375(1), 100-107. <https://doi.org/10.1016/j.canlet.2016.02.044>

## Figures

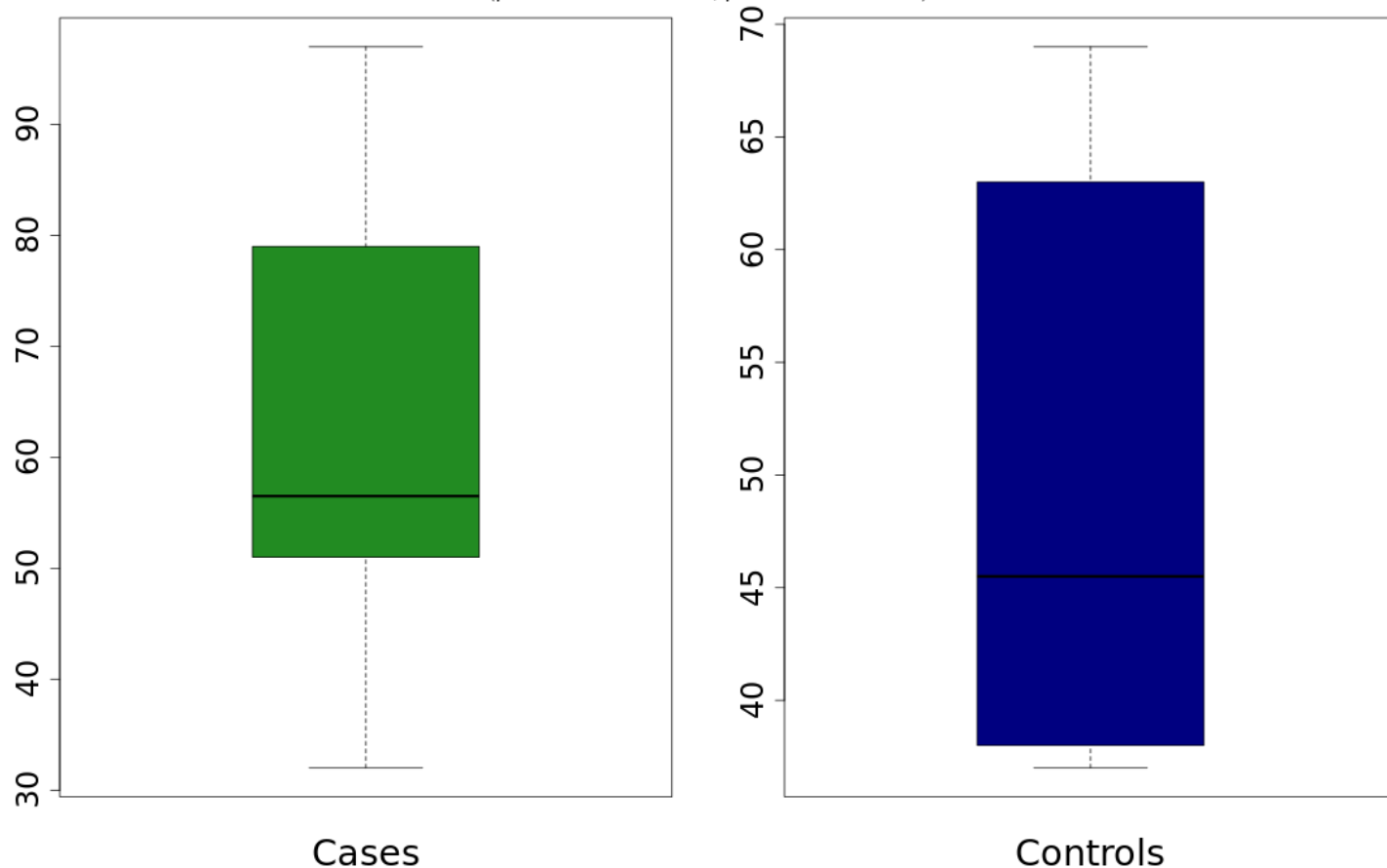


**Figure 1**

Manhattan plot of the asd2019 GWAS data. Each dot represents a gene tested for association with ASD; the dotted horizontal line represents a Bonferroni corrected p value threshold of  $2.5 \times 10^{-6}$ .

## SOX7 GE counts in GSE30573 data

( $p=2.22E-07$  ASD2019;  $p=0.0017$  ASD2017)



**Figure 2**

Boxplot demonstrating significant upregulation of SOX7 in autism cases ( $p = 0.0017$ ).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementary.docx](#)