



Article

Family-Based Genome-Wide Association Study of Autism Spectrum Disorder in Middle Eastern Families

Yasser Al-Sarraj^{1,2} , Eman Al-Dous^{1,2}, Rowaida Z. Taha², Dina Ahram^{2,3}, Fouad Alshaban² , Mohammed Tolfat⁴, Hatem El-Shanti^{2,5}  and Omar M.E. Albagha^{1,2,6,*} 

- ¹ College of Health and Life Sciences, Hamad Bin Khalifa University, Doha 34110, Qatar; yalsarraj@qf.org.qa (Y.A.-S.); eymaldous@hbku.edu.qa (E.A.-D.)
 - ² Qatar Biomedical Research Institute (QBRI), Hamad Bin Khalifa University, Doha 34110, Qatar; rtaha@hbku.edu.qa (R.Z.T.); da2747@cumc.columbia.edu (D.A.); falshaban@hbku.edu.qa (F.A.); hatem-el-shanti@uiowa.edu (H.E.-S.)
 - ³ Division of Nephrology, Columbia University Medical Center, New York, NY 10032, USA
 - ⁴ The Shafallah Center for Children with Special Needs, Doha 33123, Qatar; Mohamed.Tolefat@shafallah.org.qa
 - ⁵ Department of Pediatrics, Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA
 - ⁶ Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK
- * Correspondence: oalbagha@hbku.edu.qa

Abstract: Autism spectrum disorder (ASD) is a neurodevelopmental disease characterized by abnormalities in language and social communication with substantial clinical heterogeneity. Genetic factors play an important role in ASD with heritability estimated between 70% to 80%. Genome-wide association studies (GWAS) have identified multiple loci associated with ASD. However, most studies were performed on European populations and little is known about the genetic architecture of ASD in Middle Eastern populations. Here, we report the first GWAS of ASD in the Middle eastern population of Qatar. We analyzed **171 families** with ASD, using linear mixed models adjusting for relatedness and other confounders. Results showed that common single nucleotide polymorphisms (SNP) in seven loci are associated with ASD ($p < 1 \times 10^{-5}$). Although the identified loci did not reach genome-wide significance, many of the top associated SNPs are located within or near genes that have been implicated in ASD or related neurodevelopmental disorders. These include *GORASP2*, *GABBR2*, *ANKS6*, *THSD4*, *ERCC6L*, *ARHGEF6*, and *HDAC8*. Additionally, three of the top associated SNPs were significantly associated with gene expression. We also found evidence of association signals in two previously reported ASD-susceptibility loci (rs10099100 and rs4299400). Our results warrant further functional studies and replication to provide further insights into the genetic architecture of ASD.

Keywords: autism spectrum disorder; neuropsychiatric disorders; genetic; genome-wide association; family



Citation: Al-Sarraj, Y.; Al-Dous, E.; Taha, R.Z.; Ahram, D.; Alshaban, F.; Tolfat, M.; El-Shanti, H.; Albagha, O.M.E. Family-Based Genome-Wide Association Study of Autism Spectrum Disorder in Middle Eastern Families. *Genes* 2021, 12, 761. <https://doi.org/10.3390/genes12050761>

Academic Editor: Derek Morris

Received: 22 March 2021

Accepted: 13 May 2021

Published: 18 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Autism spectrum disorder (ASD) is characterized by aberrations in social interaction and communication that are associated with repetitive behaviors and interests, with substantial clinical heterogeneity [1]. The worldwide prevalence of ASD in children has been estimated to be approximately 1% and is about four times more common in males than females [2]. Recently, a cross-sectional survey of the Middle Eastern Qatari population between 2015 and 2018 revealed an ASD prevalence of 1.14% (95% CI: 0.89–1.46) among children aged between 6 and 11 years old [3]. There are neuropathological changes among subjects with ASD that usually occur at the early stages of brain development and influence the functional connectivity and synaptic plasticity, as well as neurotransmission [4–7]. Genetic factors play an important role in the etiology of ASD; twin studies have shown

4

that the heritability of ASD ranges from 70% to 80% [8,9]. However, ASD is highly heterogeneous in terms of clinical presentation, as well as the underlying genetic architecture. Studies using massively parallel sequencing have identified many rare de novo pathogenic variants in ASD patients [10–13]. Additionally, copy number variants (CNVs), including large deletions and duplications, have been detected across the genome in patients with ASD [14–17]. However, these variants are rare and explain a small proportion of ASD heritability, suggesting that the majority of the genetic risk of ASD could be attributed to common genetic variations [18]. Although earlier genome-wide association studies (GWAS) identified common variants and susceptibility loci, none reached genome-wide significance and a few showed consistent association [19–22]. Nonetheless, a recent large scale GWAS of about 18,000 ASD cases and 28,000 controls identified five ASD-susceptibility loci with genome-wide significance [23]. Another large-scale meta-analysis of GWAS in ASD identified a susceptibility locus on chromosome 10q24.32 [24]. However, most GWAS studies of ASD were performed on European populations [19–26], and a few on Chinese and Korean [27–29], but none in Middle Eastern populations.

Here, we report the first ASD-related GWAS in the Middle Eastern population of Qatar using a family-based approach. We identified multiple ASD susceptibility loci with suggestive evidence of association with ASD, as some were located within or near genes previously implicated in ASD or related neurodevelopmental disorders. We then investigated the functional relevance of the identified loci and showed that three loci were expression quantitative trait loci (eQTL). Additionally, we found evidence of replication signals in our data for two previously reported loci.

2. Materials and Methods

2.1. Study Subjects

The study cohort is made of 171 nuclear families (trios), each identified through a proband diagnosed with ASD by either an ADI-R (Autism Diagnostic Interview; revised) [30] or an ADOS (Autism Diagnostic Observation Schedule) [31]. Four families had additional affected siblings to the proband and forty-four families were missing one parent (incomplete trio). A signed informed consent was obtained from the participants or their legally authorized representatives. The probands had a clinical work-up that included an IQ test and other tests to identify associated comorbid conditions. All affected children were examined by a clinical geneticist (H E-S) to exclude any dysmorphic features or congenital anomalies. Due to the nature of the ascertainment source (The Shafallah Center for Children with Special Needs), all probands had associated intellectual impairment, and about 25% had epilepsy. The research project was approved by the Qatar Biomedical Research Institute's Institutional Review Board (IRB).

2.2. Genotyping

DNA was extracted from whole peripheral blood using the Gentra Puregene kit (Qiagen Sciences, Germantown, MD, USA) by following the manufacturer's guidelines. Genomic DNA quality and quantity were assessed using NanoDrop Spectrophotometer (ThermoFisher Scientific, UK). SNP genotyping was performed using Illumina Infinium Bead Chip Human1M-Duov3 (345 samples) or HumanOmniExpress-12v1-1 (150 samples) (Illumina, San Diego, CA, USA) as per manufacturer's protocol. The genotyping data and allele calling were conducted using GenomeStudio 2011.1 from Illumina.

2.3. Genome-Wide Association Analysis

Standardized quality control measures were applied to filter out low quality data from both SNP and sample levels [32] using PLINK-1.9 [33]. Genotypic data from the two genotyping arrays were merged and the analysis was focused on SNPs that were common to the two arrays ($n = 545,130$). We excluded SNPs with missing call rates $> 1\%$ ($n = 15,415$), Hardy-Weinberg equilibrium test p -value $< 1 \times 10^{-6}$ ($n = 61$), and minor allele frequency (MAF) $< 1\%$ ($n = 13,334$), leaving a total of 516,320 SNPs passing quality filters.

Furthermore, we removed individuals with missing genotype rate > 5% ($n = 1$) and those with excess heterozygosity ($n = 1$). We then conducted principal components analysis (PCA) using PLINK-1.9 [33] on a pruned set of independent ($r^2 < 0.05$) common (MAF > 1%) SNPs ($n = 58,052$) to define the population background and exclude population outliers; and excluded six samples because they were population outliers as defined by their PC1 or PC2 being more than 4 standard deviations away from sample mean values. The final set of 487 participants were included for downstream association analysis. A genomic kinship matrix of all individuals was calculated using the GenABEL package (v1.8.0) [34] implemented in R.3.4.3 [35] to correct for relatedness [34,36]. Association testing was performed using the linear mixed-effect model (LMM) by including gender and the first four population principal components as covariates using the GenABEL package [34], as described previously [37–39]. For the X chromosome, loci were coded as (0, 2) for males and (0, 1, 2) for females; the association analysis was performed separately in males and females, and association results were combined using meta-analysis as implemented in the GenABEL package. The pseudoautosomal regions of the X chromosome were not included in the analysis. Furthermore, to assess the relevance of the identified loci to ASD and related phenotypes, we used FUMA [40], PhenoScanner-v2 [41], GWAS catalog [42], and PubMed literature searches. LocusZoom was used to generate regional association plots for the identified loci [43]. The threshold for genome-wide significance was $p < 5 \times 10^{-8}$, and the threshold for suggestive evidence of association was $p < 1 \times 10^{-5}$. We then assessed loci reported in previous GWAS of ASD for replication in our dataset. First, we assessed SNPs with robust genome-wide significant association with ASD from previous studies ($p < 5 \times 10^{-8}$) [23], [24]. For SNPs that were not genotyped in our dataset, we searched for close proxies (within 100 kb) in linkage disequilibrium (LD) with the lead SNP based on LD data of European population from 1000 Genomes. Proxies were then assessed for evidence of replication in our dataset. We also assessed SNPs associated with ASD ($p < 1 \times 10^{-5}$) from the GWAS catalog [42] for replication in our dataset using “autism” or “autism spectrum disorder” as phenotypes. The binomial sign test was performed using GraphPad software (GraphPad, San Diego, CA, USA).

2.4. Polygenic Risk Score Analysis

Polygenic risk scores (PRS) were calculated using the “score” function in PLINK and were weighted by the estimated effect size of risk alleles. We assessed the performance of European-derived PRS in our cohort by calculating PRS based on loci reported by Grove et al. [23] with $p < 1 \times 10^{-6}$.

2.5. Expression Quantitative Trait Locus (eQTL) Analysis

The effect of associated SNPs on gene expression (eQTL) was assessed using the Genotype-Tissue Expression (GTEx) online database (<https://gtexportal.org/home/>, accessed on 14 April 2021). The SNP “rs” identifier was used as a search term and eQTL plots were generated for the tissue with the highest statistical significance.

3. Results

3.1. Characteristics of Study Subjects

The majority of study subjects were ASD simplex families with one child with ASD and the parents. Four families had more than one affected child. The average age (\pm standard deviation) of ASD patients was 8.4 ± 4.3 years. The ratio of male to female children with ASD was 4.6:1. The average age of the parents was 42.1 ± 8.2 years. Supplementary Table S1 shows the family structure of the study subjects. The principal component analysis showed that the majority of study subjects was located within one population cluster. We identified six subjects as population outliers, which were removed before the association analysis (Figure 1).

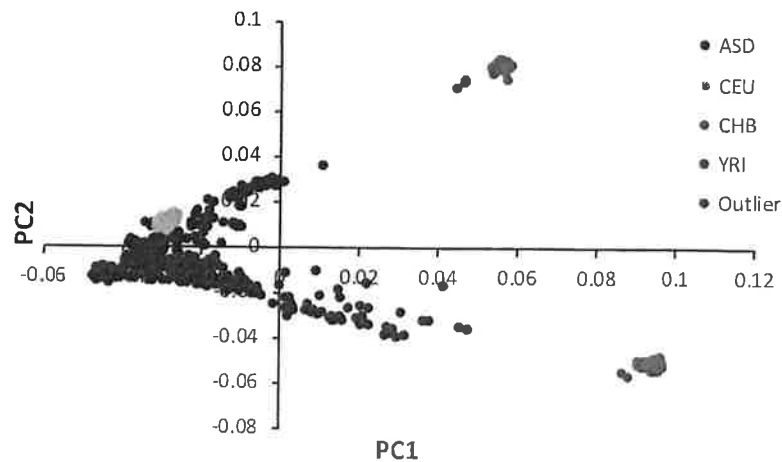


Figure 1. Population background. The first population principal component (PC1) is drawn against the second component (PC2) for study subjects (blue) as well as the three HapMap reference samples including European (CEU), East Asian (CHB) and African (YRI) populations. Population outliers were defined as samples deviating more than 4 standard deviation units from the mean population cluster and are shown in red color.

3.2. Genome-Wide Association Analysis of ASD

We used the linear mixed model to investigate the association between SNP genotypes and ASD. We included gender and the first four population PCs in the model to account for any subpopulation stratification, and the genomic kinship matrix to account for relatedness. The association testing results are summarized in the Manhattan plot (Figure 2a), which shows the top signals associated with ASD. The association testing results showed no evidence of inflation ($\lambda = 0.9997$), as shown in Figure 2b.

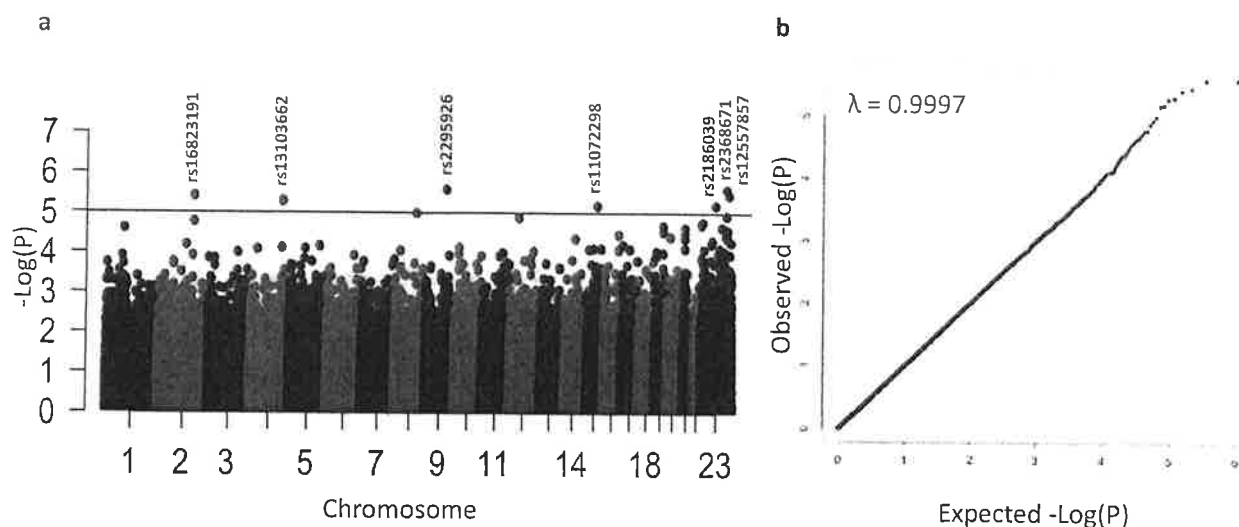


Figure 2. Manhattan and quantile-quantile plots. (a) Manhattan plot of association test results showing the chromosomal position of 516,320 analyzed SNPs plotted against $-\log(p)$. The blue horizontal line represents the threshold for suggestive evidence of association ($p < 1 \times 10^{-5}$). (b) Quantile-quantile plot showing the expected versus observed $-\log(p)$ values. The genomic inflation factor (λ) is shown at the top left corner.

Seven loci showed suggestive evidence of association ($p < 1 \times 10^{-5}$) with ASD (Table 1). The first locus is on chromosome 2q31.1 (rs16823191, $p = 3.8 \times 10^{-6}$). The associated region spans about 450 kb and contains two genes: *TLK1* and *GORASP2* (Supplementary Figure S1a). Interestingly, variants within *GORASP2* are associated with cognitive

performance in the GWAS catalog and PhenoScanner [44]. The second locus is located on chromosome 4q32.1 and tagged by rs13103662 ($p = 4.98 \times 10^{-6}$). The associated region spans about 300 kb, and is bound by two strong recombination hotspots, but does not contain mapped genes (Supplementary Figure S1b). The third locus is located on chromosome 9q22.33 and tagged by rs2295926; this locus shows suggestive evidence of association with ASD ($p = 2.57 \times 10^{-6}$; Supplementary Figure S2a). This region harbors three protein coding genes: *GABBR2*, *ANKK6*, and *GALNT12*. The fourth locus is on chromosome 15q23 with rs11072298 ($p = 6.58 \times 10^{-6}$), located within *THSD4* (Supplementary Figure S2b), in which a deleterious de novo mutation (p.P839L) is reported in ASD [11].

Table 1. Summary of SNPs suggestively associated with ASD.

SNP	Chr	Position	A1	A2	A1 Freq (%)	p	OR	95% CI	Genes
rs16823191	2	171,930,818	G	A	6.0	3.8×10^{-6}	1.22	1.09–1.36	<i>TLK1</i> , <i>GORASP2</i>
rs13103662	4	157,249,803	A	G	18.0	5.0×10^{-6}	0.89	0.83–0.95	—
rs2295926	9	101,593,825	G	A	29.5	2.6×10^{-6}	0.90	0.84–0.95	<i>GABBR2</i> , <i>ANKK6</i> , <i>GALNT12</i>
rs11072298	15	71,854,982	A	C	18.2	6.6×10^{-6}	0.89	0.83–0.95	<i>THSD4</i> <i>PIN4</i> , <i>ERCC6L</i> , <i>RPS4X</i> , <i>CITED1</i> , <i>HDAC8</i>
rs2368671	X	71,523,650	T	C	8.1	6.4×10^{-6}	1.15	1.06–1.24	<i>RPS4X</i> , <i>CITED1</i> , <i>HDAC8</i>
rs2186039	X	125,384,433	C	A	41.2	2.5×10^{-6}	1.08	1.04–1.13	<i>DCAF12L2</i>
rs12557857	X	135,868,083	G	T	2.2	3.5×10^{-6}	1.29	1.12–1.48	<i>ARHGEF6</i>

Chr, chromosome; Freq, frequency; OR, odds ratio for allele A1; CI, confidence interval. Positions are in reference to human genome build GRCh37.

The three remaining loci are located on chromosome X. The first is tagged by rs2368671 ($p = 6.38 \times 10^{-6}$) on Xq13.1 and located within *CITED1* (Supplementary Figure S3a), but this region contains other protein coding genes such as *PIN4*, *ERCC6L*, *RPS4X*, and *HDAC8*. The second is on Xq25 tagged by rs2186039 ($p = 2.55 \times 10^{-6}$), which is located close to *DCAF12L2* (Supplementary Figure S3b). *DCAF12L2* is located in a region with several reported deletions and duplications in patients with intellectual disability (ID), global developmental delay (GDD), and delayed speech and language development, as well as seizures [45]. The third is on Xq26.3, and spans about 350 kb (Supplementary Figure S4). The observed signal is tagged by rs12557857 ($p = 3.52 \times 10^{-6}$), and located within *ARHGEF6*, a gene implicated in syndromic and non-syndromic X-linked intellectual disability [46,47].

We then assessed the predictive performance of PRS derived from European populations in our cohort by calculating PRS based on loci reported by Grove et al. [23] with $p < 1 \times 10^{-6}$. Of the 466 SNPs reported by Grove et al., 24 were genotyped in our dataset and were used to calculate PRS. Results showed that European-derived PRS were not significantly associated with ASD in our cohort (OR = 1.03, 95% CI 0.96–1.11; $p = 0.41$), although subject with ASD had slightly higher PRS (21.0) compared to controls (20.8), but this was not statistically significant ($p = 0.41$; Figure 3).

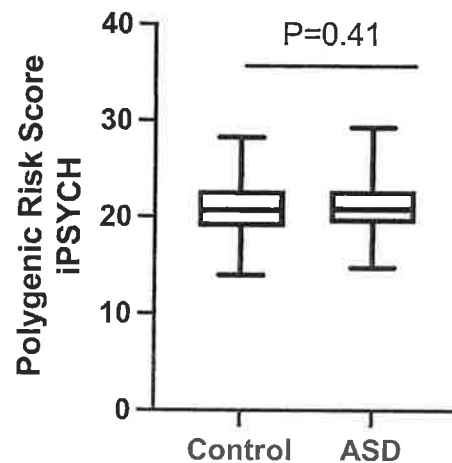


Figure 3. Analysis of polygenic risk scores (PRS). Assessment of PRS defined by SNPs with $p < 1 \times 10^{-6}$ reported by Grove et al. (iPSYCH) [23] in our cohort. Values are presented as box and whiskers plots showing the interquartile range (boxes), median (horizontal line), the minimum and maximum values (whiskers).

Since none of the suggestively associated SNPs are coding variants, we investigated their functional significance as eQTL using the Genotype-Tissue Expression (GTEx) database. We found that three of the seven top associated SNPs are significant eQTL in multiple tissues (Supplementary Figures S5–S7). eQTL plots from the tissue with highest statistical significance are shown in Figure 4. The risk allele “T” of rs2368671 is associated with reduced *PIN4* gene expression in multiple tissues (Supplementary Figure S5). Similarly, the rs2186039 allele “C” conferring risk of ASD is associated with reduced *DCAF12L2* gene expression (Figure 4 and Supplementary Figure S6) and rs12557857 genotypes are associated with *ARHGEF6* gene expression (Figure 4 and Supplementary Figure S7).

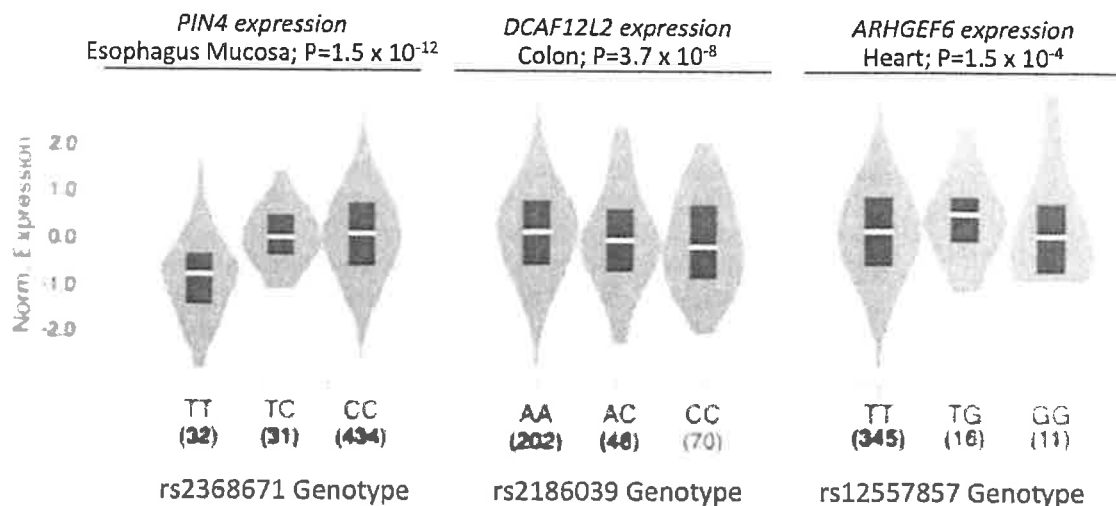


Figure 4. Gene expression in relation to genotypes of ASD-associated SNPs. Bean plots of normalized gene expression in relation to genotypes showing the median (white line) and interquartile range (black box). Data and plots were adapted from the GTEx database [48].

3.3. Replication of Loci Reported in Previous GWAS of ASD

We investigated SNPs previously reported in GWAS of ASD in our sample, focusing on variants with genome-wide significant association first. Previous large scale GWAS studies identified six SNPs with robust genome-wide significant association with ASD

($p < 5 \times 10^{-8}$) [23,24]. These include: rs201910565, rs111931861, rs10099100, rs1409313, rs71190156, and rs910805. Only one SNP (rs910805) was genotyped in our dataset, but this showed no significant association with ASD ($p > 0.05$). However, we found evidence of association in our dataset ($p < 0.05$) for multiple SNPs located close to, and in linkage disequilibrium (LD) with, two previously reported SNPs. For example, we found five SNPs in our dataset that were located within about 50 kb of rs10099100, showed evidence of association with ASD ($p < 0.05$), and were in LD with rs10099100 (Supplementary Table S2). Additionally, an SNP (rs4299400) in our dataset located within 85 kb of rs71190156 showed evidence of association with ASD ($p = 0.03$), and was in LD with rs71190156 ($D' = 0.55$). We then assessed SNPs associated with ASD ($p < 1 \times 10^{-5}$; $n = 66$) from the GWAS catalog [42] for replication in our dataset. Of the 66 SNPs from the GWAS catalog, 29 were genotyped in our dataset, but none showed significant association with ASD ($p > 0.05$; Supplementary Table S3). Comparison of the direction of effect was possible for 12 SNPs and showed a concordance rate of 58.3% (sign test p -value = 0.39).

4. Discussion

We performed a family-based, genome-wide association analysis of ASD using a cohort from the Middle East and identified seven loci with suggestive evidence for association with autism. The locus on 2q31.1 harbors two genes: *TLK1* and *GORASP2*. *TLK1* belongs to a well-conserved Tausled-like kinases (TLKs) that are involved in various cellular functions including DNA replication and repair, chromatin structure, and regulation of cell cycle. Interestingly, SNPs in *TLK1* have been associated with mathematical ability in GWAS studies [49], whereas pathogenic mutations in *TLK2*, a paralog of *TLK1*, have been reported in patients with ASD, intellectual disability (ID), schizophrenia, and microcephaly [50–53]. *GORASP2* encodes a member of the Golgi reassembly stacking protein that plays a role in Golgi ribbon formation. A de novo missense mutation in *GORASP2* has been reported in an ASD patient [54], and SNPs within this gene have been associated with cognitive performance in GWAS studies [44]. Furthermore, the 2q31.1 region has been linked to ASD and related pervasive developmental disorders in genome-wide linkage studies [55,56].

The locus on 9q22.33 harbors three genes, *GABBR2*, *ANKS6*, and *GALNT12*. *GABBR2* encodes gamma-aminobutyric acid type B receptor subunit 2, a protein that belongs to the G-protein coupled receptor and GABA-B receptor superfamily. This gene has been implicated in many neurodevelopmental and neuropsychiatric disorders [57,58]. In addition, several pathogenic variants in *GABBR2* have been reported in patients with early infantile epileptic encephalopathy [59], Rett syndrome-like disorders [60,61] and ASD [62]. On the other hand, *ANKS6* encodes a protein that contains multiple ankyrin repeats with possible role in renal and cardiovascular development. Studies have reported on two de novo variants in individuals with ASD from the Simons Simplex Collection [63]; one was intronic [64], and the other was a deleterious missense variant (p.R467Q) [11].

THSD4 is the only gene located within the 15q23 locus and has been associated, along with a combined set of genes, with “social skills” quantitative trait association analyses conducted in twins [65]. Additionally, a missense de novo variant (p.P839L) in *THSD4* has been reported in an ASD patient from the Simons Simplex Collection [63].

Two genes located within the locus on Xq13.1 have relevance to ASD. Missense de novo variants in *ERCC6L* have been reported in patients with ASD (p.Q932H) [54] and developmental disorders (p.M358T) [66]. The other is *HDAC8*, which is associated with Cornelia de Lange syndrome (CDLS5), a developmental disorder characterized by short stature and intellectual disability with variable clinical presentations [67]. Multiple missense de novo and pathogenic variants have been reported in *HDAC8* in patients with developmental disorders including those with intellectual disability [66]. There is a consensus among researchers that 20–49% of individuals with *HDAC8* pathogenic variants present with ASD [68]. A study reported a stop gain de novo pathogenic variant (p.Y174Ter) in a child with ID [69].

The *DCAF12L2* gene is located within the locus on Xq25. Large CNVs spanning this region have been reported in patients with developmental delay, intellectual disability, and ASD [70]. The locus on Xq26.3 harbors many genes, but the strongest association signal was located within *ARHGEF6*. This region is linked to X-linked mental retardation [62]. Moreover, mutations in *ARHGEF6* have been reported in patients with X-linked mental retardation [47] and in multiple families with ASD [11,12,71,72].

Although the association signals did not reach genome-wide significance ($p < 5 \times 10^{-8}$) in our study, many were located within genes that have been implicated in ASD or related neurodevelopmental disorders.

Moreover, three of the top associated SNPs were significantly associated with gene expression suggesting a regulatory function. We also found evidence of association signals in two previously reported, genome-wide significant, ASD susceptibility loci (rs10099100 and rs4299400) suggesting common genetic architecture across populations for these two loci. However, we found no evidence of replication in our dataset for SNPs previously associated with ASD at $p < 1 \times 10^{-5}$ from the GWAS catalog. This is consistent with a previous meta-analysis of 14 ASD cohorts (totaling 7387 ASD subject and 8567 control) [24], which did not identify any genome-wide significant locus, and few signals were consistent among the 14 different studies. We found that European-derived PRS were not significantly associated with ASD in our cohort. Despite being consistent with previous reports that showed PRS derived from European populations have lower predictive performance when applied to middle eastern populations [37], proper evaluation of PRS requires further studies with larger sample sizes and higher number of SNPs. In our study, we were only able to evaluate 24 out of the 466 SNPs reported by Grove et al. [23].

A limitation of this study is the small sample size. However, the family-based approach utilized in this study is less likely to be influenced by population stratifications compared to population-based case-control designs. Another limitation of this study is the lack of genotype imputation, but this was not performed due to the absence of appropriate reference haplotype panels for the Middle Eastern population to allow accurate imputation of genotypes. Our data warrant further functional studies and replication to confirm their association with ASD and provide further insights into the genetic architecture of ASD.

5. Conclusions

This study has identified suggestive evidence of association between ASD and common SNPs in several genes and loci that are suspected of playing a role in ASD or related neurodevelopmental disorders. Our results are consistent with previous studies which showed that common genetic variations in multiple loci contribute to ASD-susceptibility and that the genetic architecture of ASD is complex and shared with other neurodevelopmental disorders. This study has provided new insights into the genetic architecture of ASD in the Middle East.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/genes12050761/s1>, Supplementary Table S1: Family structure of the study subjects, Supplementary Table S2: Replication signals for the lead SNP rs10099100, Supplementary Table S3: Replication analysis for SNPs with $p < 1 \times 10^{-5}$ from the GWAS catalog, Supplementary Figure S1: Regional association plot for chr 2q31.1 (a) and 4q32.1 (b) regions, Supplementary Figure S2: Regional association plot for chr 9q22.33 (a) and 15q23 (b) regions, Supplementary Figure S3: Regional association plot for chr Xq13.1 (a) and Xq25 (b) regions, Supplementary Figure S4: Regional association plot for chromosome Xq26.3 region, Supplementary Figure S5: *PIN4* expression in multiple tissues in relation to rs2368671 genotype, Supplementary Figure S6: *DCAF12L2* expression in multiple tissues in relation to rs2186039 genotype, Supplementary Figure S7: *ARHGEF6* expression in multiple tissues in relation to rs12557857 genotype.

Author Contributions: O.M.E.A. designed and supervised the study; Y.A.-S., E.A.-D., R.Z.T. and D.A. performed sample processing and phenotype quality control; Y.A.-S. conducted data analysis; O.M.E.A. and Y.A.-S. wrote the manuscript with a contribution from H.E.-S.; H.E.-S. lead study subject recruitment and phenotyping; F.A., M.T. and Members of The Shafallah Center contributed

to subject recruitment, phenotyping, and data acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by start-up grants to O.M.E.A. from the college of health and life sciences and the Qatar Biomedical Research Institute at Hamad Bin Khalifa University. Y.A.-S. and E.A.-D. are supported by a PhD scholarship from Hamad Bin Khalifa University.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of Qatar Biomedical Research Institute (2010-002).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: All data supporting the findings of this study are available either within the article, the supplementary data, or from the authors upon reasonable request.

Acknowledgments: We thank the patients and their families who participated in this study for their infinite assistance and persistence.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

References

1. American Psychiatric Association Division of Research. Highlights of Changes from DSM-IV to DSM-5. *Focus* **2013**, *11*, 525–527. [CrossRef]
2. Levy, S.E.; Mandell, D.S.; Schultz, R.T. Autism. *Lancet* **2009**, *374*, 1627–1638. [CrossRef]
3. Alshaban, F.; Aldosari, M.; Al-Shammari, H.; El-Hag, S.; Ghazal, I.; Tolefat, M.; Ali, M.; Kamal, M.; Aati, N.A.; Abeidah, M.; et al. Prevalence and correlates of autism spectrum disorder in Qatar: A national study. *J. Child Psychol. Psychiatry* **2019**, *60*, 1254–1268. [CrossRef] [PubMed]
4. Ross, P.J.; Zhang, W.-B.; Mok, R.S.; Zaslavsky, K.; Deneault, E.; D’Abate, L.; Rodrigues, D.C.; Yuen, R.K.; Faheem, M.; Mufteev, M.; et al. Synaptic Dysfunction in Human Neurons With Autism-Associated Deletions in PTCHD1-AS. *Biol. Psychiatry* **2020**, *87*, 139–149. [CrossRef]
5. Guang, S.; Pang, N.; Deng, X.; Yang, L.; He, F.; Wu, L.; Chen, C.; Yin, F.; Peng, J. Synaptopathology Involved in Autism Spectrum Disorder. *Front. Cell. Neurosci.* **2018**, *12*, 470. [CrossRef] [PubMed]
6. Shen, M.D.; Li, D.D.; Keown, C.L.; Lee, A.; Johnson, R.T.; Angkustsiri, K.; Rogers, S.J.; Müller, R.-A.; Amaral, D.G.; Nordahl, C.W. Functional Connectivity of the Amygdala Is Disrupted in Preschool-Aged Children With Autism Spectrum Disorder. *J. Am. Acad. Child Adolesc. Psychiatry* **2016**, *55*, 817–824. [CrossRef] [PubMed]
7. Olivito, G.; Clausi, S.; Laghi, F.; Tedesco, A.M.; Baiocco, R.; Mastropasqua, C.; Molinari, M.; Cercignani, M.; Bozzali, M.; Leggio, M. Resting-State Functional Connectivity Changes Between Dentate Nucleus and Cortical Social Brain Regions in Autism Spectrum Disorders. *Cerebellum* **2016**, *16*, 283–292. [CrossRef] [PubMed]
8. Bailey, A.; Le Couteur, A.; Gottesman, I.; Bolton, P.; Simonoff, E.; Yuzda, E.; Rutter, M. Autism as a strongly genetic disorder: Evidence from a British twin study. *Psychol. Med.* **1995**, *25*, 63–77. [CrossRef]
9. Rosenberg, R.E.; Law, J.K.; Yenokyan, G.; McGready, J.; Kaufmann, W.E.; Law, P.A. Characteristics and Concordance of Autism Spectrum Disorders Among 277 Twin Pairs. *Arch. Pediatr. Adolesc. Med.* **2009**, *163*, 907–914. [CrossRef]
10. Sadybekov, A.; Tian, C.; Arnesano, C.; Katritch, V.; Herring, B.E. An autism spectrum disorder-related de novo mutation hotspot discovered in the GEF1 domain of Trio. *Nat. Commun.* **2017**, *8*, 1–13. [CrossRef]
11. Turner, T.N.; Coe, B.P.; Dickel, D.E.; Hoekzema, K.; Nelson, B.J.; Zody, M.C.; Kronenberg, Z.N.; Hormozdiari, F.; Raja, A.; Pennacchio, L.A.; et al. Genomic Patterns of De Novo Mutation in Simplex Autism. *Cell* **2017**, *171*, 710–722.e12. [CrossRef]
12. Yuen, R.K.C.; Merico, D.; Bookman, M.; Howe, J.L.; Thiruvahindrapuram, B.; Patel, R.V.; Whitney, J.; Deflaux, N.; Bingham, J.; Wang, Z.; et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat. Neurosci.* **2017**, *20*, 602–611. [CrossRef]
13. An, J.-Y.; Lin, K.; Zhu, L.; Werling, D.M.; Dong, S.; Brand, H.; Wang, H.Z.; Zhao, X.; Schwartz, G.B.; Collins, R.L.; et al. Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder. *Science* **2018**, *362*, eaat6576. [CrossRef]
14. Glessner, J.T.; Wang, K.; Cai, G.; Korvatska, O.; Kim, C.E.; Wood, S.; Zhang, H.; Estes, A.; Brune, C.W.; Bradfield, J.P.; et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* **2009**, *459*, 569–573. [CrossRef]
15. Pinto, D.; Pagnamenta, A.T.; Klei, L.; Anney, R.; Merico, D.; Regan, R.; Conroy, J.; Magalhaes, T.R.; Correia, C.; Abrahams, B.S.; et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **2010**, *466*, 368–372. [CrossRef]
16. O’Roak, B.J.; Stessman, H.A.; Boyle, E.A.; Witherspoon, K.T.; Martin, B.; Lee, C.; Vives, L.; Baker, C.; Hiatt, J.B.; Nickerson, D.A.; et al. Recurrent de novo mutations implicate novel genes underlying simplex autism risk. *Nat. Commun.* **2014**, *5*, 1–6. [CrossRef] [PubMed]

17. Sebat, J.; Lakshmi, B.; Malhotra, D.; Troge, J.; Lese-Martin, C.; Walsh, T.; Yamrom, B.; Yoon, S.; Krasnitz, A.; Kendall, J.; et al. Strong Association of De Novo Copy Number Mutations with Autism. *Science* **2007**, *316*, 445–449. [CrossRef]
18. Gaugler, T.; Klei, L.; Sanders, S.J.; Bodea, C.A.; Goldberg, A.P.; Lee, A.B.; Mahajan, M.C.; Manaa, D.; Pawitan, Y.; Reichert, J.G.; et al. Most genetic risk for autism resides with common variation. *Nat. Genet.* **2014**, *46*, 881–885. [CrossRef]
19. Wang, K.; Zhang, H.; Ma, D.; Bucan, M.; Glessner, J.T.; Abrahams, B.S.; Salyakina, D.; Imielinski, M.; Bradfield, J.P.; Sleiman, P.M.A.; et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nat. Cell Biol.* **2009**, *459*, 528–533. [CrossRef] [PubMed]
20. Ma, D.; Salyakina, D.; Jaworski, J.M.; Konidari, I.; Whitehead, P.L.; Andersen, A.N.; Hoffman, J.D.; Slifer, S.H.; Hedges, D.J.; Cukier, H.N.; et al. A Genome-wide Association Study of Autism Reveals a Common Novel Risk Locus at 5p14.1. *Ann. Hum. Genet.* **2009**, *73*, 263–273. [CrossRef]
21. Weiss, L.A.; The Gene Discovery Project of Johns Hopkins & the Autism Consortium; Arking, D.E. A genome-wide linkage and association scan reveals novel loci for autism. *Nat. Cell Biol.* **2009**, *461*, 802–808. [CrossRef]
22. Anney, R.; Klei, L.; Pinto, D.; Regan, R.; Conroy, J.; Magalhaes, T.R.; Correia, C.; Abrahams, B.S.; Sykes, N.; Pagnamenta, A.T.; et al. A genome-wide scan for common alleles affecting risk for autism. *Hum. Mol. Genet.* **2010**, *19*, 4072–4082. [CrossRef] [PubMed]
23. Grove, J.; Ripke, S.; Als, T.D.; Mattheisen, M.; Walters, R.K.; Won, H.; Pallesen, J.; Agerbo, E.; Andreassen, O.A.; Anney, R.; et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **2019**, *51*, 431–444. [CrossRef]
24. The Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. 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. *Mol. Autism* **2017**, *8*, 1–17. [CrossRef]
25. Connolly, J.J.; Glessner, J.T.; Hakonarson, H. A Genome-Wide Association Study of Autism Incorporating Autism Diagnostic Interview-Revised, Autism Diagnostic Observation Schedule, and Social Responsiveness Scale. *Child Dev.* **2012**, *84*, 17–33. [CrossRef]
26. Anney, R.; Klei, L.; Pinto, D.; Almeida, J.; Bacchelli, E.; Baird, G.; Bolshakova, N.; Bölte, S.; Bolton, P.F.; Bourgeron, T.; et al. Individual common variants exert weak effects on the risk for autism spectrum disorders. *Hum. Mol. Genet.* **2012**, *21*, 4781–4792. [CrossRef] [PubMed]
27. Xia, L.; Ou, J.; Li, K.; Guo, H.; Hu, Z.; Bai, T.; Zhao, J.; Xia, K.; Zhang, F. Genome-wide association analysis of autism identified multiple loci that have been reported as strong signals for neuropsychiatric disorders. *Autism Res.* **2020**, *13*, 382–396. [CrossRef] [PubMed]
28. Kuo, P.-H.; Chuang, L.-C.; Su, M.-H.; Chen, C.-H.; Chen, C.-H.; Wu, J.-Y.; Yen, C.-J.; Wu, Y.-Y.; Liu, S.-K.; Chou, M.-C.; et al. Genome-Wide Association Study for Autism Spectrum Disorder in Taiwanese Han Population. *PLoS ONE* **2015**, *10*, e0138695. [CrossRef]
29. Cho, S.-C.; Yoo, H.J.; Park, M.; Cho, I.H.; Kim, B.-N.; Kim, J.-W.; Shin, M.-S.; Park, T.-W.; Son, J.-W.; Chung, U.-S.; et al. Genome-Wide Association Scan of Korean Autism Spectrum Disorders with Language Delay: A Preliminary Study. *Psychiatry Investig.* **2011**, *8*, 61–66. [CrossRef]
30. Lord, C.; Rutter, M.; Le Couteur, A. Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Disord.* **1994**, *24*, 659–685. [CrossRef] [PubMed]
31. Lord, C.; Risi, S.; Lambrecht, L.; Cook, E.H., Jr.; Leventhal, B.L.; DiLavore, P.C.; Pickles, A.; Rutter, M. The Autism Diagnostic Observation Schedule—Generic: A Standard Measure of Social and Communication Deficits Associated with the Spectrum of Autism. *J. Autism Dev. Disord.* **2000**, *30*, 205–223. [CrossRef] [PubMed]
32. Laurie, C.C.; Doherty, K.F.; Mirel, D.B.; Pugh, E.W.; Bierut, L.J.; Bhargava, T.; Boehm, F.; Caporaso, N.E.; Cornelis, M.C.; Edenberg, H.J.; et al. Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet. Epidemiology* **2010**, *34*, 591–602. [CrossRef]
33. Chang, C.C.; Chow, C.C.; Tellier, L.C.A.M.; Vattikuti, S.; Purcell, S.M.; Lee, J.J. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* **2015**, *4*, 7. [CrossRef]
34. Aulchenko, Y.S.; Ripke, S.; Isaacs, A.; Van Duijn, C.M. GenABEL: An R library for genome-wide association analysis. *Bioinformatics* **2007**, *23*, 1294–1296. [CrossRef]
35. R Core Team. *A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2016; Volume 2.
36. Thomson, R.; McWhirter, R. Adjusting for Familial Relatedness in the Analysis of GWAS Data. In *Methods in Molecular Biology*; Humana Press: New York, NY, USA, 2017; Volume 1526, pp. 175–190.
37. Thareja, G.; The Qatar Genome Program Research (QGPR) Consortium; Al-Sarraj, Y.; Belkadi, A.; Almotawa, M.; Suhre, K.; Albagha, O.M.E. Whole genome sequencing in the Middle Eastern Qatari population identifies genetic associations with 45 clinically relevant traits. *Nat. Commun.* **2021**, *12*, 1–10. [CrossRef]
38. Fakiola, M.; Strange, A.; Cordell, H.J.; Miller, E.N.; Pirinen, M.; Su, Z.; Mishra, A.; Mehrotra, S.; Monteiro, G.R.; Band, G.; et al. Common variants in the HLA-DRB1–HLA-DQA1 HLA class II region are associated with susceptibility to visceral leishmaniasis. *Nat. Genet.* **2013**, *45*, 208–213. [CrossRef] [PubMed]
39. Eu-Ahsunthornwattana, J.; Howey, R.A.; Cordell, H.J. Accounting for relatedness in family-based association studies: Application to Genetic Analysis Workshop 18 data. *BMC Proc.* **2014**, *8*, S79. [CrossRef]

40. Watanabe, K.; Taskesen, E.; Van Bochoven, A.; Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **2017**, *8*, 1–11. [CrossRef] [PubMed]
41. Staley, J.R.; Blackshaw, J.; Kamat, M.A.; Ellis, S.; Surendran, P.; Sun, B.B.; Paul, D.S.; Freitag, D.; Burgess, S.; Danesh, J.; et al. PhenoScanner: A database of human genotype–phenotype associations. *Bioinformatics* **2016**, *32*, 3207–3209. [CrossRef]
42. MacArthur, J.; Bowler, E.; Cerezo, M.; Gil, L.; Hall, P.; Hastings, E.; Junkins, H.; McMahon, A.; Milano, A.; Morales, J.; et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **2017**, *45*, D896–D901. [CrossRef]
43. Pruim, R.J.; Welch, R.P.; Sanna, S.; Teslovich, T.M.; Chines, P.S.; Gliedt, T.P.; Boehnke, M.; Abecasis, G.R.; Willer, C.J. LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics* **2010**, *26*, 2336–2337. [CrossRef] [PubMed]
44. Need, A.C.; Attix, D.K.; McEvoy, J.M.; Cirulli, E.T.; Linney, K.L.; Hunt, P.; Ge, D.; Heinzen, E.L.; Maia, J.M.; Shianna, K.V.; et al. A genome-wide study of common SNPs and CNVs in cognitive performance in the CANTAB. *Hum. Mol. Genet.* **2009**, *18*, 4650–4661. [CrossRef] [PubMed]
45. Firth, H.V.; Richards, S.M.; Bevan, A.P.; Clayton, S.; Corpas, M.; Rajan, D.; Van Vooren, S.; Moreau, Y.; Pettett, R.M.; Carter, N.P. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am. J. Hum. Genet.* **2009**, *84*, 524–533. [CrossRef] [PubMed]
46. Yntema, H.G.; Hamel, B.C.; Smits, A.P.; Van Roosmalen, T.; Helm, B.V.D.; Kremer, H.; Ropers, H.H.; Smeets, D.F.; Van Bokhoven, H. Localisation of a gene for non-specific X linked mental retardation (MRX46) to Xq25-q26. *J. Med. Genet.* **1998**, *35*, 801–805. [CrossRef] [PubMed]
47. Kutsche, K.; Yntema, H.; Brandt, A.; Jantke, I.; Nothwang, H.G.; Orth, U.; Boavida, M.G.; David, D.; Chelly, J.; Fryns, J.-P.; et al. Mutations in ARHGEF6, encoding a guanine nucleotide exchange factor for Rho GTPases, in patients with X-linked mental retardation. *Nat. Genet.* **2000**, *26*, 247–250. [CrossRef]
48. Lonsdale, J.; Thomas, J.; Salvatore, M.; Phillips, R.; Lo, E.; Shad, S.; Hasz, R.; Walters, G.; Garcia, F.; Young, N.; et al. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **2013**, *45*, 580–585. [CrossRef]
49. Lee, J.J.; Wedow, R.; Okbay, A.; Kong, E.; Maghzian, O.; Zacher, M.; Nguyen-Viet, T.A.; Bowers, P.; Sidorenko, J.; Linnér, R.K.; et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **2018**, *50*, 1112–1121. [CrossRef]
50. Segura-Bayona, S.; Stracker, T.H. The Tousel-like kinases regulate genome and epigenome stability: Implications in development and disease. *Cell. Mol. Life Sci.* **2019**, *76*, 3827–3841. [CrossRef]
51. Lelieveld, S.H.; Reijnders, M.R.F.; Pfundt, R.; Yntema, H.G.; Kamsteeg, E.-J.; De Vries, P.; De Vries, B.B.A.; Willemsen, M.H.; Kleefstra, T.; Löhner, K.; et al. Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nat. Neurosci.* **2016**, *19*, 1194–1196. [CrossRef]
52. Reijnders, M.R.; Miller, K.A.; Alvi, M.; Goos, J.A.; Lees, M.M.; De Burca, A.; Henderson, A.; Kraus, A.; Mikat, B.; De Vries, B.B.; et al. De Novo and Inherited Loss-of-Function Variants in TLK2: Clinical and Genotype-Phenotype Evaluation of a Distinct Neurodevelopmental Disorder. *Am. J. Hum. Genet.* **2018**, *102*, 1195–1203. [CrossRef]
53. Fromer, M.; Pocklington, A.J.; Kavanagh, D.H.; Williams, H.J.; Dwyer, S.; Gormley, P.; Georgieva, L.; Rees, E.; Palta, P.; Ruderfer, D.M.; et al. De novo mutations in schizophrenia implicate synaptic networks. *Nat. Cell Biol.* **2014**, *506*, 179–184. [CrossRef]
54. Iossifov, I.; O’Roak, B.J.; Sanders, S.J.; Ronemus, M.; Krumm, N.; Levy, D.; Stessman, H.A.; Witherspoon, K.T.; Vives, L.; Patterson, K.E.; et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **2014**, *515*, 216–221. [CrossRef]
55. Lauritsen, M.B.; Als, T.D.; A Dahl, H.; Flint, T.J.; Wang, A.G.; Vang, M.; A Kruse, T.; Ewald, H.; Mors, O. A genome-wide search for alleles and haplotypes associated with autism and related pervasive developmental disorders on the Faroe Islands. *Mol. Psychiatry* **2005**, *11*, 37–46. [CrossRef]
56. De Bruijn, D.; Van Dijk, A.; Pfundt, R.; Hoischen, A.; Merkx, G.; Gradek, G.; Lybæk, H.; Stray-Pedersen, A.; Brunner, H.; Houge, G. Severe Progressive Autism Associated with Two de novo Changes: A 2.6-Mb 2q31.1 Deletion and a Balanced t(14;21)(q21.1;p11.2) Translocation with Long-Range Epigenetic Silencing of LRFN5 Expression. *Mol. Syndr.* **2010**, *1*, 46–57. [CrossRef]
57. Fatemi, S.H.; Folsom, T.D.; Thuras, P.D. Deficits in GABAB receptor system in schizophrenia and mood disorders: A postmortem study. *Schizophr. Res.* **2011**, *128*, 37–43. [CrossRef]
58. Hedges, D.J.; Hamilton-Nelson, K.L.; Sacharow, S.J.; Nations, L.; Beecham, G.W.; Kozhekbaeva, Z.M.; Butler, B.L.; Cukier, H.N.; Whitehead, P.L.; Ma, D.; et al. Evidence of novel fine-scale structural variation at autism spectrum disorder candidate loci. *Mol. Autism* **2012**, *3*, 2. [CrossRef]
59. Hamdan, F.F.; Myers, C.T.; Cossette, P.; Lemay, P.; Spiegelman, D.; Laporte, A.D.; Nassif, C.; Diallo, O.; Monlong, J.; Cadieux-Dion, M.; et al. High Rate of Recurrent De Novo Mutations in Developmental and Epileptic Encephalopathies. *Am. J. Hum. Genet.* **2017**, *101*, 664–685. [CrossRef]
60. Yoo, Y.; Jung, J.; Lee, Y.-N.; Lee, Y.; Cho, H.; Na Bs, E.; Hong, J.; Kim, E.; Lee, J.S.; Lee, J.S.; et al. GABBR2 mutations determine phenotype in rett syndrome and epileptic encephalopathy. *Ann. Neurol.* **2017**, *82*, 466–478. [CrossRef]
61. Lopes, F.; Barbosa, M.; Ameer, A.; Soares, G.; De Sá, J.; Dias, A.I.; Oliveira, G.; Cabral, P.; Temudo, T.; Calado, E.; et al. Identification of novel genetic causes of Rett syndrome-like phenotypes. *J. Med. Genet.* **2016**, *53*, 190–199. [CrossRef]
62. Takata, A.; Miyake, N.; Tsurusaki, Y.; Fukai, R.; Miyatake, S.; Koshimizu, E.; Kushima, I.; Okada, T.; Morikawa, M.; Uno, Y.; et al. Integrative Analyses of De Novo Mutations Provide Deeper Biological Insights into Autism Spectrum Disorder. *Cell Rep.* **2018**, *22*, 734–747. [CrossRef]

63. Fischbach, G.D.; Lord, C. The Simons Simplex Collection: A Resource for Identification of Autism Genetic Risk Factors. *Neuron* **2010**, *68*, 192–195. [CrossRef] [PubMed]
64. Michaelson, J.J.; Shi, Y.; Gujral, M.; Zheng, H.; Malhotra, D.; Jin, X.; Jian, M.; Liu, G.; Greer, D.; Bhandari, A.; et al. Whole-Genome Sequencing in Autism Identifies Hot Spots for De Novo Germline Mutation. *Cell* **2012**, *151*, 1431–1442. [CrossRef]
65. Hu, V.W.; Devlin, C.A.; Debski, J.J. ASD Phenotype—Genotype Associations in Concordant and Discordant Monozygotic and Dizygotic Twins Stratified by Severity of Autistic Traits. *Int. J. Mol. Sci.* **2019**, *20*, 3804. [CrossRef]
66. Study, D.D.D. Deciphering Developmental Disorders Study Prevalence and architecture of de novo mutations in developmental disorders. *Nature* **2017**, *542*, 433–438. [CrossRef]
67. Harakalova, M.; Boogaard, M.-J.V.D.; Sinke, R.; Van Lieshout, S.; Van Tuil, M.C.; Duran, K.; Renkens, I.; Terhal, P.A.; De Kovel, C.; Nijman, I.J.; et al. X-exome sequencing identifies aHDAC8variant in a large pedigree with X-linked intellectual disability, truncal obesity, gynaecomastia, hypogonadism and unusual face. *J. Med. Genet.* **2012**, *49*, 539–543. [CrossRef]
68. Kline, A.D.; Moss, J.F.; Selicorni, A.; Bisgaard, A.-M.; Deardorff, M.A.; Gillett, P.M.; Ishman, S.L.; Kerr, L.M.; Levin, A.V.; Mulder, P.A.; et al. Diagnosis and management of Cornelia de Lange syndrome: First international consensus statement. *Nat. Rev. Genet.* **2018**, *19*, 649–666. [CrossRef] [PubMed]
69. Msc, P.B.; Joset, P.; Steindl, K.; Oneda, B.; Gogoll, L.; Azzarello-Burri, S.; Sheth, F.; Datar, C.; Verma, I.C.; Puri, R.D.; et al. Elucidation of the phenotypic spectrum and genetic landscape in primary and secondary microcephaly. *Genet. Med.* **2019**, *21*, 2043–2058. [CrossRef]
70. Kaminsky, E.B.; Kaul, V.; Paschall, J.; Church, D.M.; Bunke, B.; Kunig, D.; Moreno-De-Luca, D.; Moreno-De-Luca, A.; Mülle, J.G.; Warren, S.T.; et al. An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. *Genet. Med.* **2011**, *13*, 777–784. [CrossRef]
71. O’Roak, B.J.; Vives, L.; Girirajan, S.; Karakoc, E.; Krumm, N.; Coe, B.P.; Levy, R.; Ko, A.; Lee, C.; Smith, J.A.B.; et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **2012**, *485*, 246–250. [CrossRef]
72. Jónsson, H.; Sulem, P.; Kehr, B.; Kristmundsdóttir, S.; Zink, F.; Hjartarson, E.; Hardarson, M.T.; Hjorleifsson, K.E.; Eggertsson, H.P.; Gudjonsson, S.A.; et al. Parental influence on human germline de novo mutations in 1548 trios from Iceland. *Nature* **2017**, *549*, 519–522. [CrossRef]