Exploring autism and autism co-occurring condition associations to elucidate multivariate genetic mechanisms and insights

Karoliina Salenius[1]*, Niina Väljä [1]*, Sini Thusberg[1], *The GEMMA study group*[2], Christine Ladd-Acosta[3], Matti Nykter[1,4], Christophe Roos[5], Alessio Fasano[6,7], Reija Autio[8]*, Jake Lin [8,9]*
*Equal contributions

Affiliations
1. Faculty of Medicine and Health Technology, Tampere University and Tays Cancer Centre, Tampere, Finland
2. The GEMMA study, GEMMA-PROJECT.eu, Salerno, Italy
3. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
4. Foundation for the Finnish Cancer Institute, Helsinki, Finland
5. Euformatics, Tekniikantie, Espoo, Finland
6. European Biomedical Research Institute of Salerno (EBRIS), Salerno, Italy
7. Harvard Medical School, Harvard T.H. Chan School of Public Health, Boston, USA
8. Health Sciences, Faculty of Social Sciences, Tampere University, Tampere, Finland
9. Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden

Abstract:

Background: Genome-wide association (GWAS) and large scaled exome based studies have successfully uncovered hundreds of associations and genes related to autism. Autism is a partially heritable neurodevelopmental condition. Autistic people may also have other co-occurring conditions such as ADHD, anxiety disorders, depression, mental health issues, learning difficulty, physical health conditions and communication challenges. The concomitant development of autism and other neurological conditions is assumed to result from a complex interplay between genetics and the environment. However, only a limited number of studies have performed multivariate genetic autism association studies.

Methods: We conducted the largest multivariate GWAS with autism and 11 autism co-occurring traits (ADHD, ADHD child, anxiety, IBS, depression, digestion disorders, education attainment,

epilepsy, panic, social phobia) using summary statistics from leading reputable consortiums: iPSYCH, PGC, 23andMe and UK Biobank. Multivariate associations and central traits were further identified. Subsequently, colocalization and Mendelian randomization (MR) analysis were performed on the associations identified with the central traits containing autism. To further validate our findings we utilized independent datasets consisting of 35,584 exome sequence iPSYCH data and 53 (13 probands) Whole Genome Sequence from the GEMMA project.

Results: Multivariate GWAS resulted in 380 significant associations (p-value < 5e-8), among which 203 are reported for the first time for any trait. 22 SNPs were identified to contain autism and one or more comorbidities in their central trait set, including variants mapped to known SFARI autism genes such as MAPT and NEGR1. NSF, a serotonin receptor binding gene, was found to be associated with autism with education attainment. Mendelian randomization and colocalization analyses found that education attainment, ADHD and ADHD childhood are liable traits toward autism. Additionally, allele proportions and distribution differences between MAPT (17q21.31) region aberrations and MAPT H1/H2 haplotypes, known to associate with neurodevelopment, and the reported autism-central SNPs were found within GEMMA autism probands and controls. Pathway and gene card analysis also implicated microbiome, enteric inflammation, and central nervous system enrichments.

Conclusions: Our study combined multivariate genome-wide association testing with systematic decomposition to identify novel genetic associations related to autism and autism co-occurring driver traits. Statistical tests were applied to discern evidence for shared and interpretable liability between autism and co-occurring traits. We additionally detected differences in the SNP association frequency patterns between autism probands and familial controls in an independent validation cohort. These findings contribute to our understanding of the complex genetic underpinnings of autism and provide valuable insights into its potential driver traits and genetic markers.

Introduction

Autism spectrum condition (autism) is an umbrella term for a group of heterogeneous neurodevelopmental conditions that manifest in early childhood. Autism is associated with polygenic markers and with several key factors such as family history, genetics and environment

(Khachadourian et al. 2023; Nayar et al. 2021; Chaste and Leboyer 2012). The diagnosis of autism is based its key characteristics including difficulties in social communication and interaction, restricted and repetitive behaviors, hyperactivity and divergent responses to sensory inputs. The most common co-occurring conditions in autistic persons are attention deficit hyperactivity disorder (ADHD), ADHD childhood, anxiety, depression, epilepsy, obsessive compulsive disorders (OCD), panic attacks, social phobia and general social phobia, all of which share overlapping diagnostic attributes and challenging symptoms with autism (Khachadourian et al. 2023; Romero et al. 2016). According to US data, autistic children tend to fare less well in educational attainment (EA) and about one in three have a reduced intellectual ability, as defined by intelligence quotient (IQ less than 70) (Baio et al. 2018; Tamm et al. 2020). Some autistic children with higher IQ scores also comparatively experience harder academic struggles due to co-occurring conditions and difficulties in social interactions (Ashburner, Ziviani, and Rodger 2010).

Together with recent advances in genomics technology and pivotal support from the engaged autism community, 1,162 human genes are currently implicated with autism development and these are curated in the SFARI (Arpi and Simpson 2022; Grove et al. 2019; Pinto et al. 2014) gene module. These curated genes, with varying degrees of effect, are scored using the Evaluation of Autism Gene Link Evidence (EAGLE) framework (Schaaf et al. 2020). Surprisingly, while it is known that common variants contribute to the majority of genetic background (Gaugler et al. 2014), only a few robust genetic associations have been recently reported. Most of these are attributed to the landmark study conducted by Grove and colleagues, employing a large Danish cohort with 18,381 autism cases and 27,969 controls, where 12 significant variant associations were reported (Grove et al. 2019).

Given that there is overlap in symptoms between autism and ADHD, a genetics study recently found shared genetic factors underlying autism and ADHD (Peyre et al. 2021), with partial concordance between bidirectional colocalization single nucleotide variants (SNPs). However, the study was limited to general ADHD (onset age 10+), and not childhood ADHD.

Astoundingly many (47% median) autistic children have reported one or more gastrointestinal (GI) symptoms (Holingue et al. 2018; Boorstein 2008). Recently, there have been promising results that link microbiome disruption and diversity (Morton et al. 2023) as a novel contributing factor to

autism. Interestingly, while Grove and colleagues found that 7 of the 12 autism SNP associations have similar significance towards EA and psychosis traits depression and schizophrenia (Grove et al. 2019), very little is known concerning the joint and the shared genetic mechanisms between autism and autism co-occurring traits including ADHD, ADHD chilhood, EA, depression and gastrointestinal symptoms.

To address these knowledge gaps in autism and in exploration of potential shared co-occurring trait genetic associations, this study performed multivariate genome wide association, using available summary statistics from large reputable cohorts, to explore novel multivariate genetic associations. We aimed to extract the statistically significant autism biomarkers driving the multivariate associations. To achieve this, colocalization was systematically applied to test the robustness between the shared variants and conditions (Wallace 2021). In addition, mendelian randomisation was further applied, using the multivariate variants and the essential biomarkers, to assess liability relationships between autism and the selected co-occurring conditions (Bowden, Davey Smith, and Burgess 2015; Rees, Wood, and Burgess 2017). We performed functional enrichment with our multivariate associations as well as validated our findings relative to recent studies on putative genes and their associations to autism development (Satterstrom et al. 2020; Feliciano et al. 2019).

Materials and methods

Genome-wide summary statistics for autism and ADHD were collected from the PGC and iPSYCH (Pedersen et al. 2018; Sullivan et al. 2018) studies. Education attainment (Okbay et al. 2022) summary file was collected from the Social Science Genetic Association Consortium (SSGAC). Additional autism co-occurring conditions were obtained from the UK Biobank (Neale lab) project as well as iPSYCH. The 12 studies, with sample sizes ranging from 46,351-765,283 are shown in detail in Table 1 below. Summary statistics are joined, yielding 4,525,476 SNPs, and applied in a multivariate GWAS setting. Systematic post analysis includes decomposition aiming to detect the most important traits and subsequent colocalization and Mendelian randomisation analysis to explore shared liability and interpretable mechanisms, as shown in Figure 1 workflow. Additional method and implementation details are captured below.

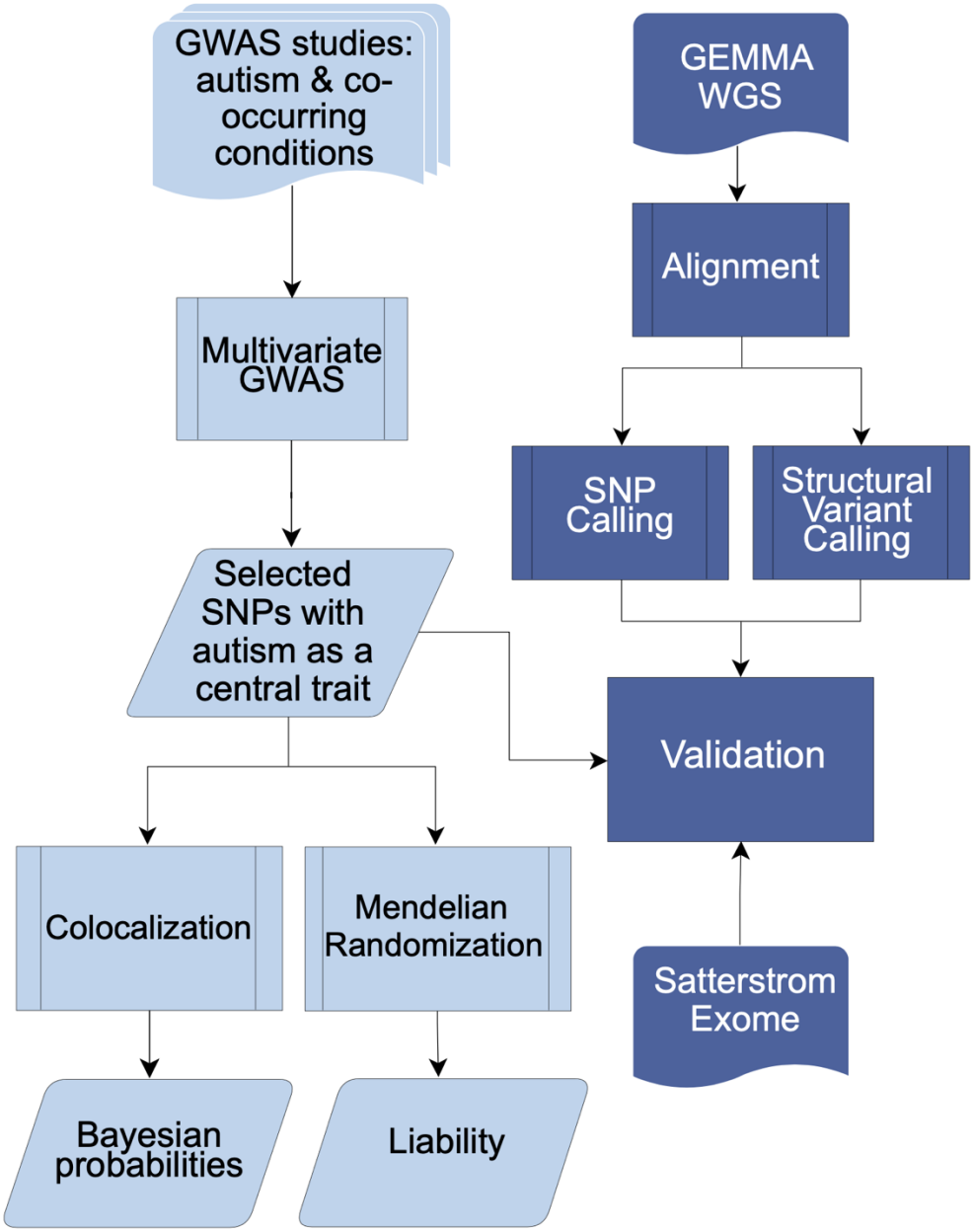


Figure 1: Workflow for the analyses conducted in the study. Multivariate GWAS was performed on selected GWAS studies including autism and 11 co-occurring traits: ADHD, ADHD childhood, anxiety, depression, digestion, educational attainment, obsessive compulsive disorder, panic attack, social phobia, irritable bowel syndrome and epilepsy. 22 SNPs were selected and evaluated with Colocalization and Mendelian Randomization. Further validation of the selected SNPs utilized independent datasets from Satterstrom et.al (2019) as well as the GEMMA -study. The GEMMA whole genome sequencing (WGS) processing included variant calling to infer structural and single nucleotide variants (SVs and SNVs) present in the samples.

Multivariate GWAS and determination of central traits

Multivariate GWAS on autism and autism co-occurring biomarker traits were performed using metaPhat/metaCCA software that performs multivariate analysis by implementing Canonical Correlation Analysis (CCA) for a set of univariate GWAS summary statistics (Ruotsalainen et al. 2021; Cichonska et al. 2016; Lin et al. 2020). The objective of metaCCA is to find the optimal linear combination of the predictor variables that is maximally correlated with a linear combination of the response variables. Autism multivariate central traits, are identified by MetaPhat decomposition based on iterative tracing and testing, and are defined as the subset of driver traits that are essential to the calculation of the significant p-value (without these traits, the joint p-value exceeds 5e-8). The central traits also include the subset of traits yielding the lowest Bayesian Information Criterion (BIC), representing the best model fit. At each iteration, the biomarker to be excluded is the one whose exclusion leads to the highest p-value for the remaining subset. Similarly, Bayesian information criterion (BIC) values, a metric for model fit (Schwarz 1978), are iteratively decomposed. The driver traits are determined as a set of traits that have been removed when the multivariate p-value becomes non-significant ($p > 5e-8$). The central traits are the union of the driver biomarker traits with the subset yielding the lowest BIC value. The interpretation is that these central biomarker traits make the multivariate association significant and approximately optimal.

Colocalization analyses

We performed colocalization analyses for the selected multivariate autism SNPs to assess if the associated variants in the locus are shared (Supplementary table 2) genetically between autism and co-occurring related autism traits to account for erroneous results that may follow from analyzing individual SNPs. Errors can occur when a SNP associated with trait 1 and trait 2 are in linkage disequilibrium (LD). Another instance is when one SNP is associated with both traits, causing the less strongly associated trait to be missed altogether. The colocalization analyses were performed using the R package coloc (Giambartolomei et al. 2014; Wallace 2020).

The colocalization analysis was conducted using the absolute base factor colocalization method (coloc.abf), which is a Bayesian colocalization analysis method. It uses Bayes Factors to calculate posterior probabilities for different variant combinations assuming there is a single liable variant

161 for each trait. A relatively large region size window of 100KB (50±KB flanking the SNP position)
162 was selected to comprehensively span large genes (Piovesan et al. 2016) and cover all potential
163 LD and recombination elements. The different hypotheses for which posterior probabilities are
164 calculated include: H0 (no liable variant), H1 (liable variant only for trait 1), H2 (liable variant
165 only for trait 2), H3 (two separate liable variants, no colocalazion e.g. two individual SNPs in
166 LD), H4 (common liable variant shared between the traits, with colocalization). As recommended
167 (Wallace 2020), default setting prior probability thresholds were applied for different hypotheses:
168 1e-4 for H1, H2 and H3 and 1e-5 for H4. The SNPs that colocalize (H4) between autism and EA
169 are considered as shared.
170

171 ## Mendelian randomization analyses

172 We conducted Mendelian Randomization analyses (MR) on the selected autism multi-trait SNPs
173 based on their assigned central traits, to explore the liability and independent (reverse causation)
174 relationships between autism and its related biomarkers (Phillips and Smith 1991). Similar to
175 coloc.abf analysis, a regional window of 100KB (50±KB flanking) was applied for the selected
176 variants within their relevant trait summary statistics. For those variants inside a gene locus the
177 window was also set to the gene length plus 25±KB, to conservatively cover the median gene
178 length of 26 KB (Piovesan et al. 2016) as well as covering regulatory and promoter elements within
179 its flanking regions. The analyses were performed using the R package MendelianRandomization
180 (version 0.9, Bowden, Davey Smith, and Burgess 2015).
181

182 For variants with two central traits, both inverse variance weighted (IVW) and MR-Egger -
183 methods were utilized (Bowden, Davey Smith, and Burgess 2015). The IVW-method, based on
184 Cochran's Q statistical test for heterogeneity (Cochran 1941), uses the assigned inverse weighted
185 effect sizes to estimate the liable effect of the exposure to the outcome, so that those variants with
186 smaller standard errors of the effect size get assigned more weight. To control for the effect of
187 horizontal pleiotropy the MR-analysis was additionally conducted using the MR-Egger method
188 with testing the condition of the correlation between genetic associations with the exposure and
189 direct effects on the outcome to be approximately zero (Instrument Strength Independent of Direct
190 Effect (Bowden, Davey Smith, and Burgess 2015)). Multivariate IVW -method (Rees, Wood, and

191 Burgess 2017) was used for those variants with more than two central traits. Multivariate IVW-
192 analysis was implemented using the multivariate weighted linear regression option. The
193 significance threshold used to calculate the confidence interval was set to 0.05 for all methods
194 except genome-wide association tests where 5e-8 was applied (Risch and Merikangas 1996).

195 Genetic annotations, pathway enrichment and validation

196 SFARI Base Gene resource and Open targets were used to assess the novelty of variants and genes
197 in relation to autism development. To estimate pathway enrichment, Reactome and Elsevier
198 databases pathway enrichments were evaluated with the Enrichr (Kuleshov et al. 2016) tool.
199 Human organ systems and their tissue cell type enrichment analysis, encompassing 1,355 tissue-
200 cell type (TS) panel, with an additional 111 single-cell RNAseq data, was conducted using
201 WebCSEA (Dai et al. 2022, Lake et al. 2018). Autism association burden validation was further
202 performed on exome sequencing data from Satterstrom and colleagues (Satterstrom et al. 2020).

203 Whole genome sequencing

204 We further validated our results using yet unpublished data from the EU Horizon2020 GEMMA
205 research project. We used genotype variant calls in 53 (13 proband) whole genome sequencing
206 samples from the GEMMA prospective cohort (Troisi et al. 2020). Fifty of these samples were
207 sourced from 15 Italian families with one additional US family providing three more samples.
208 These samples, assayed on whole blood and collected during enrollment, were sequenced with
209 30X coverage on Illumina NovaSeq 6000 platform as paired end short reads (2x150bp). Each
210 sample contained on average 993M high quality reads. Samples were aligned to GRCh38 reference
211 genome using bwa mem version 0.7.17 (Li 2013) and aligned reads were sorted and marked for
212 duplicates with samtools version 1.12 (Li et al. 2009). The average read depth after alignment was
213 47. Quality control at all data processing steps was performed with omnomicsQ -software
214 (Euformatics, Finland; Gutowska-Ding et al. 2020). For variant calling we utilized DeepVariant
215 version 1.4.0 (Poplin et al. 2018) after which samples were genotyped using GLnexus version
216 1.4.1 (Yun et al. 2021, https://github.com/dnanexus-rnd/GLnexus) configured for DeepVariant,
217 resulting in ~4M variants per sample. Small variants were loaded on the omnomicsNGS software
218 (Euformatics, Finland) for an individual and cohort-based first analysis. Structural variants were
219 called with Manta software version 1.6.0 (Chen et al. 2015), which was run family-wise with

220 default parameters. The resulting structural variants from the families were merged to a single
221 multisample vcf -file with Truvari version 4.4.0 collapse -tool using the default parameters
222 (English et al. 2022). A 900kb region on 17:45922635-46822634 was explored for breakends
223 reported by Manta, indicating the presence of an inversion. Coordinates for the 238bp CNV in
224 MAPT gene intron 9 were obtained by visualizing the structural variant data in IGV version 2.12.3.

225 Statistical analysis

226 All statistical analyses were performed using R 4.2.2 software and available as R markdown results
227 in our github project. Genome-wide association is called on the standard and strict p-value
228 threshold of 5e-8, to account for multiple testing based on the assumption of about 1-million
229 independent tests (Risch and Merikangas 1996). Bonferroni correction is applied for multiple
230 testing, including within Multivariable Mendelian randomisation (MVMR) (Yavorska and
231 Burgess 2017) results.
232

233 Results

234 GWAS summary statistics

235 Genome-wide summary statistics for autism and ADHD were collected from the PGC and
236 iPSYCH (Pedersen et al. 2018; Sullivan et al. 2018) studies. Education attainment (Okbay et al.
237 2022) summary file was collected from the Social Science Genetic Association Consortium
238 (SSGAC). Additional autism co-occurring conditions were obtained from the UK Biobank (Neale
239 lab) project as well as iPSYCH (Supplementary table 1). We note that Schizophrenia is not
240 incorporated from the UK Biobank due to the low number (157) of cases. Study sample sizes and
241 number of SNPs along with the genetic correlation computed from 4,525,476 SNPs (found in all
242 summaries) are shown in Table 1. In line with previous studies (Grove et al. 2019), the largest
243 genetic correlation values between autism and co-occurring traits were with ADHD (0.264) and
244 ADHD childhood (0.260), followed by correlation with educational attainment (0.040)
245 (Supplementary Figure 1).
246
247

248

249 Table 1. Summary statistics from autism and autism related genome-wide association studies are applied to explore
250 autism genetic and trait refinement.
251

Trait	GWAS Source	N (Cases)	N SNPs	Genetic correlation with ASD
Autism	iPSYCH+PGC	46,351 (18,382)	9,112,386	NA
ADHD	iPSYCH+PGC	225,534 (38,691)	6,774,224	0.264
ADHD Childhood	iPSYCH	60,736 (15,338)	5,795,968	0.260
Educational Attainment	SSGAC/23andMe	765,283 (EA Years nonbinary/mean 15.74)	10,985,948	0.040
OCD	Psychiatric Genomics Consortium (PGC)	9,725 (2,688)	13,791,467	0.001
Depression	UKB Neale lab	361,194 (33,398)	13,791,467	-0.022
Epilepsy	UKB Neale lab	361,194 (8,990)	13,791,467	-0.003
General Anxiety	UKB Neale lab	361,194 (22,017)	13,791,467	-0.016
GI symptoms/Other digestion issues	UKB Neale lab	361,194 (27,424)	13,791,467	0.008
Irritable Bowel Syndrome (IBS)	UKB Neale lab	361,194 (36,219)	13,791,467	0.014
Panic Attack	UKB Neale lab	361,194 (8,695)	13,791,467	-0.011
Social Phobia	UKB Neale lab	361,194 (1,959)	13,791,467	-0.001

252

Multivariate autism central trait SNPs, pathway and organ tissue enrichment

Using GWAS summary statistics from autism and autism co-occurring traits (shown in Table 1) ADHD, ADHD childhood, anxiety, depression, digestion disorder, education attainment, epilepsy, IBS, OCD, panic attack and social phobia from iPSYCH, UKB and 23andMe, we found 380 (p-value < 5e-8) SNP associations, including 203 variants that are reported for the first time for any trait (Supplementary Table 6). While none of the 380 SNPs mapped to the common autism genetic effect variants of autism study (Grove et al. 2019), when assessed at the gene level, 18 genes were concordant (as indicated in STable 6). Decomposition implemented in MetaPhat, using stepwise tracing of p-value and Bayesian information criteria (BIC) contributions, was applied to identify autism central trait sets and 22 SNPs remained where 10 have been previously linked to existing GWAS studies but not autism (p-value of 12 trait multi-GWAS and p-value of decomposed central traits test are listed in Supplementary Table 7 and trace plots available in Supplementary data) (Lin et al. 2020; Schwarz 1978). These multivariate autism central SNPs mapped to 22 genes (Table 2) where three: CADPS, MAPT and NEGR1 are known SFARI genes, with autism susceptibility EAGLE curated scores >= 2 (indicating lower liability evidence due to limited experimental results) (Schaaf et al. 2020). SNP variant rs538628 is found within the NSF gene, a receptor for serotonin (Iwata et al. 2014) and known marker of autism severity, is also reported. Educational attainment, ADHD and ADHD childhood, as shown in Supplementary Figure 2, were the most common co-autism -central traits. P-value and BIC decomposition processing of MAPT and NSF are shown in Supplementary Figure 3.

Shown in Supplementary Table 8, Figure 2e and Supplementary Figure 4, pathway enrichment using the 22 genes mapped from the 22 SNPs was performed with Enrichr (Kuleshov et al. 2016) and as expected, found neural and microtubule structural related pathway hits from Reactome (Fabregat et al. 2018) and Elsevier pathway (Nesterova et al. 2019) database resources. The top pathways were Synaptic Exocytosis in Dementia (MAPT and NSF, p-value 0.00026), Genes with Mutations Associated with Frontotemporal Dementia (MAPT, p-value 0.006) and Hypoxia related pathways (BNIP3, p-value 0.024). We additionally performed human organ tissue system enrichment analysis using WebCSEA (Dai et al. 2022), using the selected 22 central autism trait gene associations and importantly found significance (Figure 2f, p-value < 1e-03) with the digestive, nervous, and sensory organ systems. As shown in Supplementary Figure 5, the most

enriched tissue types are related to cerebrum, cortex, and intestine related resources discerned from 1,355 tissue-type (TS) as well as data from the human brain single cell project (Lake et al. 2018). Furthermore, filtering on gene card (Safran et al. 2010) curations and listed in Supplementary Table S5, we additionally found that these mapped genes towards annotations related to gut, microbiome, intestinal immune, enteric nervous and central nervous systems. The reported autism co-occurring MAPT and NSF associations serve as key validations for the role of microbiome-driven neuroinflammatory and impairment of gut barrier implicated in the development of autism as well as Alzheimer's disease (Fiorentino et al. 2016, Aaldijk et al. 2022).

The variant rs538628 within the NSF gene, a known serotonin transporter (Iwata et al. 2014), previously only implicated in mice with autism-like behaviors (Xie et al. 2021), is identified to associate with the central trait set of autism and education attainment. The improved autism central trait set p-values, relative to the GWAS with 12 traits, are listed in Supplementary Table S6.

Table 2: Multivariate GWAS autism-central SNPs tested with coloc and Mendelian randomisation tests towards the identified autism central traits, with 5/22 SNPs passing both as indicated with x. The order of the central traits are determined by p-value importance during decomposition processing. SFARI ASD gene members are indicated with asterisks.

RSID	Gene	Chrom:Pos ref>alt	Central Traits	coloc	MR
rs6699841	NEGR1*	1:72645850A>G	Autism, EA, ADHD	x	x
rs10932536	ENSG00000197585	2:215369762T>C	Autism, ADHD, ADHDCHILD, EA		x
rs1455350	ENSG00000225421	2:199497115A>T	Autism ADHDCHILD EA		x
rs76076331	PDIA6	2:10977585C>T	Autism EA		
rs6707462	TSGA10	2:99648689C>T	Autism ADHD EA		
rs62243489	CADPS*	3:62482927T>G	Autism EA		x
rs6773869	PLCL2	3:16870340A>G	Autism EA		
rs228619	MANBA	4:103569283A>G	Autism ADHD ADHDCHILD		x
rs406413	ENSG00000246316	5:113898581A>T	Autism EA		x
rs2400170	ENST00000659832	5:144545930C>A	Autism ADHD ADHDCHILD		x
rs9375188	ENSG00000271860	6:98555272C>T	Autism EA	x	
rs1982564	ENSG00000254115	8:87776287T>G	Autism EA		x
rs4735931	ERICH1-AS1	8:751823C>T	Autism ADHD ADHDCHILD	x	x

rs2379268	ENST00000673759	8:142640905G>A	Autism EA		
rs12761761	BNIP3	10:133775375C>T	Autism EA		
rs2237944	SERGEF	11:17838116C>T	Autism EA		
rs10149787	NPC2	14:74925452G>T	Autism EA ADHD		x
rs2732706	ARL17B	17:44351686C>T	Autism EA	x	x
rs434805	CRHR1-IT1	17:43718336T>C	Autism EA		x
rs62061734	MAPT*	17:44018488T>C	Autism EA		x
rs538628	NSF	17:44787313G>C	Autism EA	x	x
rs809220	RN7SKP140	20:21414037T>C	Autism ADHD ADHDCHILD	x	x

Colocalization analyses

Colocalization analysis was conducted on the 22 multivariate SNP associations identified to contain autism as a central trait. The comparative analysis was performed on the relevant mapped gene window, from start to end while adding 25 KBs on both ends to cover all potential regulating and promoter regional elements. For the three SNPs that did not map to a gene, the window size used for the colocalization analysis was 100 KB (50 KB from SNP position on both ends), estimated and derived from the median length of human gene of 26KB. Additional information concerning the number of regional LD adjusted SNPs applied to the colocalization test is shown in Supplementary Table 2.

We found that a total of 6/22 SNPs showed strong evidence for a common liability variant with autism in the locus (passed H4 (common liable variant shared between autism and selected central trait) with probability > 0.8, details shown in Supplementary Table 3). The traits having common liable variants inside the locus included ADHD with 8/22 as a central trait passing. Childhood ADHD had 3/6 associations passing while 4/18 education attainment linked associations passed, including SNP rs538628 that aligns with gene NSF (N-ethylmaleimide sensitive factor) the probability of a shared liable variant between autism and EA is 100 % (p-value1e-05) in the given locus (Supplementary Table 3). Colocalisation processing using the original summary statistics of autism and educational attainement for rs62061734 (MAPT, failed), and rs538628 (NSF, passed), depicting supporting regional SNPs (X axis) and their p-value (neg log10, Y axis) and effect direction (triangles as negative) are shown in Figure 2. Colocalisation failed for MAPT due to a proximal unselected SNP (rs62056859 BP 43994435)

with a slightly smaller p-value (8.86e-21) compared to the selected SNP rs62061734 with p-value of 9.62e-21, resulting in a failed H4 probability of 8.19%.

Mendelian randomization analyses

Mendelian randomization analysis was conducted for 22 SNPs. For 15 of these SNPs the central traits were significantly positively associated with autism (Table 2). The central traits which were found to be associated with a statistically significant (p-value < 0.001, accounting for multiple testing with Bonferroni correction of 12 traits) increased probability of autism were ADHD childhood, ADHD and EA, with significant positive association cases of all cases being 75 %, 70 % and 45 %, respectively (Supplementary table 4). For these cases horizontal pleiotropy did not have a significant effect on the results verified with the MR-Egger method.

The Multivariate Mendelian Randomization results for SNPs inside the window of rs538628 associated gene NSF (50±KB) suggest an increased probability of autism associated with EA (Figure 2c). The IVW-method provides a strong effect estimate of 3.63 (95% CI: 3.33-3.93) p-value < 1e-3, which is in accordance with the MR-Egger analysis results (3.88 (95% CI: 2.95-4.81), p-value < 1e-3). The MR-Egger method did not detect any significant horizontal pleiotropy (β-intercept = -0.004 (standard error 0.008), p-value = 0.581). Similarly, for MAPT rs62061734 SNP (Figure 2d), the IVW-method between autism and EA provided an effect estimate of 3.88 (95% CI: 3.83-3.94) with p-value < 1e-5, similar to the MR-Egger estimates of 3.92 (95% CI: 3.72-4.11), further supporting a strong and independent, as genotypes are fixed at conception thus minimizing reverse causation, association between EA and autism (Haycock et al. 2016).

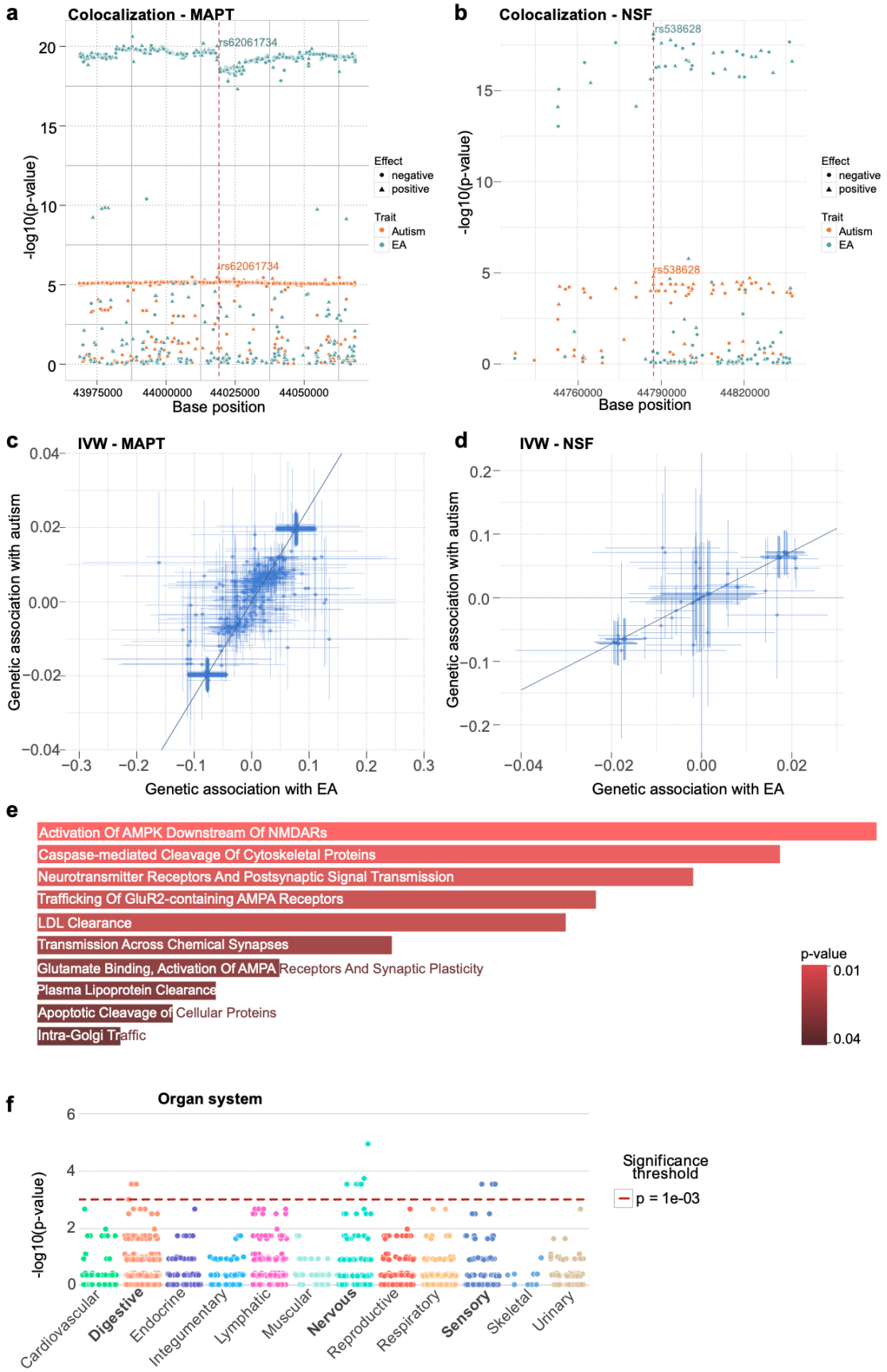

a
Colocalization - MAPT
rs62061734
rs62061734
-log10(p-value)
Base position
Effect
negative
positive
Trait
Autism
EA
b
Colocalization - NSF
rs538628
rs538628
-log10(p-value)
Base position
Effect
negative
positive
Trait
Autism
EA
c
IVW - MAPT
Genetic association with autism
Genetic association with EA
d
IVW - NSF
Genetic association with autism
Genetic association with EA
e
Activation Of AMPK Downstream Of NMDARs
Caspase-mediated Cleavage Of Cytoskeletal Proteins
Neurotransmitter Receptors And Postsynaptic Signal Transmission
Trafficking Of GluR2-containing AMPA Receptors
LDL Clearance
Transmission Across Chemical Synapses
Glutamate Binding, Activation Of AMPA Receptors And Synaptic Plasticity
Plasma Lipoprotein Clearance
Apoptotic Cleavage of Cellular Proteins
Intra-Golgi Traffic
p-value
0.01
0.04
f
Organ system
-log10(p-value)
Significance threshold
p = 1e-03
Cardiovascular
Digestive
Endocrine
Integumentary
Lymphatic
Muscular
Nervous
Reproductive
Respiratory
Sensory
Skeletal
Urinary

Figure 2: Results from the post GWAS analysis of the 22 selected SNPs. **a,b)** Colocalization processing using the original summary statistics of autism and EA for (a) rs62061734 (MAPT, failed colocalization with H4 probability 8.19%, p-value 0.09), and (b) rs538628 (NSF, passed colocalization with H4 probability 100%, p-value 1.1e-05), depicting supporting regional SNPs (x-axis) and their negative log10 p-value (y-axis) and effect direction (circles negative, triangles positive). **c,d)** Mendelian randomization (MR) results using inverse variance weighted (IVW) - method for association of autism SNP effects (y-axis) and EA SNP effects (x-axis) in the NSF (c) and MAPT (d) genes. **e)** Pathway analysis for the genes associated with the selected SNPs show enrichment in processes related to neurons using Reactome database. The length of the bar represents the significance of that specific gene-set or pathway and the color indicates the significance of the pathway. Details of the pathways and genes with their associated p-values are listed in Supplementary Table 8. **f)** Organ system enrichment was applied using WebCSEA, using the selected 22 multivariate gene associations and found enrichment (p-value < 1e-03) with the autism relevant digestive, nervous and sensory organ systems.

Validation

Using associations identified from a large-scale exome sequencing study (Satterstrom et al. 2020), we found higher concordant proportions between the autism-central trait related SNPs and the affected exome samples, 104/11974 (.9%) compared to 20/3574 (.5%) within the unaffected sibling genome. SNP concordance is counted only on high (genotype quality >= 25) confidence associations and includes an adjustment of a window of 500KB. We note that associations of MAPT and NSF genes are among the affected autism hits in the concordant SNPs.

Additionally, we test and explore the distribution of these Autism-central trait related SNPs in 53 (13 autism proband) family based GEMMA (Troisi et al. 2020) WGS samples. Visualized in Figure 3 and listed in Supplementary Table 9, we found distinct clustering and proportional differences between the 13 proband and 40 controls in the described neurological developmental genes MAPT and NSF. MAPT SNP was present in 3/13 (23%) compared to 18/40 (45%) while for NSF probands, the same 3/13 (23%) and enriched in 17/40 (42.5%) of the controls. Separated by approximately 770K bases apart, these two genes are both found in the 17q21.31 arm and referred to as the MAPT H1/H2 haplotype region. We further overlay the autism multivariate-GWAS variants in the context of the Grove et al. (Grove et al. 2019) study as a Manhattan plot (Supplementary Figure 4), capturing ARL17B, MAPT and NSF. The MAPT H1 haplotype, associated with the Tau microtubule-associated protein, has been previously implicated in

Parkinson's disease and other neurological development disorders (Wang and Mandelkow 2016; Wider et al. 2010). We confirmed that the current reported MAPT and NSF autism linked SNPs are enriched with the MAPT H2 haplotype (shown in Figure 3), as well as breakpoint and deletion events, thus impacting MAPT alternate splicing and potentially reduction aggregation of Tau (Corsi et al. 2022).

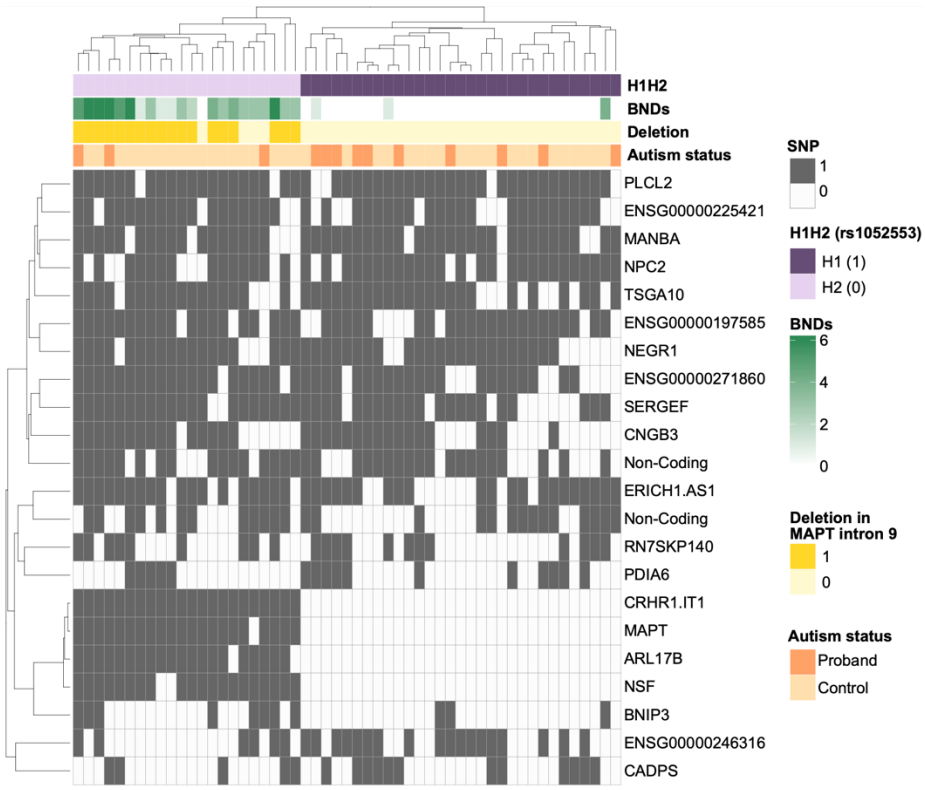


Figure 3: Results from the validation with GEMMA -cohort. Heatmap shows the clustering of the selected SNPs in the cohort, with rows showing the genes associated to each SNP and samples shown in columns. Samples with structural variation in chr17q21.31 can be seen to cluster together on the left. Green annotation visualizes the number of breakends (BNDs) detected in the 900KB region, which are suggestive of an inversion. Darker yellow annotation highlights the samples with the 238bp deletion located in intron 9 of MAPT -gene, also associated with the H2 haplotype. Assessment of the H1/H2 haplotype for MAPT was further evaluated using the SNP rs1052553. H1 haplotype has been shown to increase risk of other neurological disorders such as Altzheimer and Parkinson's Disease, and as seen in the GEMMA cohort, autism probands tend to express the H1 haplotype in darker purple (10 out of 13).

Discussion

Using multivariate statistical learning approaches, we comprehensively performed the largest autism related multi-trait GWAS analysis with summary statistics of autism, ADHD, ADHD childhood, anxiety, depression, digestion disorder, EA, epilepsy, IBS, OCD, panic, social phobia

402 to explore and decomposed the complex relationships between the genetics of autism and its most
403 common co-occurring traits. We found 22 autism driver trait sets (12 novel) genetic associations,
404 previously mapped to central and gut neuron developmental as well as inflammation and
405 microbiome domains, further underscoring the intersection of genome and microbiome as well as
406 supportive of the gut-brain axis hypothesis driving autism development(Morton et al. 2023; Cheng
407 et al. 2023). Our genetic associations pertaining to MAPT, NSF and ARL17B and microtubule
408 dysregulation and cytoskeleton pathways (Supplementary Table 8) adds additional support to the
409 canon of impairments in the regulation of the brain-barrier and gut permeability underpinning au-
410 tism development (Bonfiglio et al. 2021l; Fiorentino et al. 2016). Using the multivariate autism
411 driver trait gene sets, based on comprehensive human tissue cell type and single cell data (Dai et
412 al. 2022, Lake et al. 2018) analysis, we detected enrichment associating with digestive, nervous,
413 and sensory organ systems (Figure 2f). At the tissue cell type level and further supporting the gut-
414 brain axis and blood brain barrier, the analysis detected enriched autism relevant signals related to
415 fetal cerebrum and small intestine panels as well as frontal and visual cortex collected from a single
416 cell brain project (Lake et al. 2018).
417 Upon systematic decomposition analysis of the multivariate ASD associations, the identified au-
418 tism driver trait sets were particularly suitable for downstream analysis with shared colocalisation
419 and autism onset liability via Mendelian randomisation applications. Central autism biomarkers
420 can also be applied to assist with the construction of existing psychiatric based autism polygenic
421 scores (Jansen et al. 2020), as demonstrated for other multi-genetic complex conditions such as
422 coronary heart disease and type 2 diabetes (Lin et al. 2023; Tamlander et al. 2022). These
423 multivariate autism annotations are mapped to genes including MAPT and NSF that are known
424 members of biological pathways driving neural disorders such as infantile epilepsy (Suzuki et al.
425 2019), Parkinson's Disease (PD) (Brion, Octave, and Couck 1994; Derkinderen et al. 2021) and
426 serotonin transport (Iwata et al. 2014). A MAPT association (rs17649553) has recently been cited
427 as a strong (p-value 4.86x10-37) association, with protective effect towards PD (Nalls et al. 2014).
428 BNIP3, CADPS and NEGR1 genes have been implicated in autism microbiome features (Cheng
429 et al. 2023) and behavior developments (Singh et al. 2019). BNIP3 also has been identified to
430 associate with abnormalities in mitochondrial synaptic cells (Rojas-Charry et al. 2021) and its
431 variant rs12761761, is a known positive association of higher intelligence score (Savage et al.

2018), further confirmed in our results with a significant and negative Mendelian randomisation test effect between EA and autism.

As part of validation, overall clustering and distribution proportion differences of the identified autism central SNP associations were detected between probands and non-autistic subjects from testing on independent genomes from WGS samples collected in the GEMMA (Troisi et al. 2020) prospective cohort. As the MAPT haplotype region (17q21.31) has been linked to neural disorder development and general cognitive functions (Trampush et al. 2017), a deeper analysis revealed that deletion and structural aberration events were highly depleted in autism probands (77%, 10/13). Deletion and structural events are known to disrupt alternate splicing and thus decrease potential for the aggregation of tau, which leads to instability and structural impairment in the brain neuronal cells (Avila et al. 2004). We also found proband and control proportional differences were likewise confirmed within the evaluation of autism putative de novo variants resource reported in the large exome sequencing study from Satterstrom and colleagues (Satterstrom et al. 2020).

The most common co-central driving traits in our set of 22 multivariate associations along with autism were EA, ADHD and ADHD childhood. As it is known that the diagnosis for autism and ADHD, particularly ADHD manifestation in young children, is similar with symptomatic issues concerning hyperactivity and attention span (Kern et al. 2015). A previous study have performed systematic comparison of genetic and functional enrichment between associations unique and common to autism and ADHD (Peyre et al. 2021) GWAS resources, we further compliment their results by inclusion of other autism co-occurring traits, including ADHD and ADHD childhood as well as EA. Interestingly, autism and ADHD have been linked to disruption in microbiome composition and function (dysbiosis), gastrointestinal and bowel habits issues (Morton et al. 2023). Notably, we found variant rs2732706 mapping to ARL17B gene, have previously been only associated with a GWAS on the stool frequency trait and here is associated with autism and EA, and further exploration of the BIC and p-value decomposition traces (Lin et al. 2020) reveals that general anxiety and digestion problems contributed the 3rd and 4th (besides autism and EA) most to the BIC and p-value.

Our validation results are limited by the relatively small sample size currently available in GEMMA. Nevertheless, the deep sequencing data has allowed us to infer interesting observations concerning distribution of proband autism-central trait associations compared to familial controls. The higher enrichment of proband distribution of the reported multivariate associations are also confirmed by the analysis performed on Satterstrom and colleagues study of exome sequencing in the neurobiology of autism (Satterstrom et al. 2020). The future release of other omics data types from autism cohorts and foundation, including additional transcriptome, microbiome, metabolomics and methylation, will allow for improved statistical power with deeper temporal analysis towards confirmation of gut-brain-axis changes and genetic patterns driving autism heterogeneity and development.

Conclusion

Our study, representing the largest autism multi-trait GWAS, combined 12 genome-wide association summaries and subsequently performed systematic decomposition to identify genetic associations related to autism and autism co-occurring driver traits. Upon analysis of colocalisation and Mendelian randomisation, EA, ADHD and ADHD childhood were found to be strong independent associations towards autism with gene mappings and enriched organ and brain tissues implicated in neural development and gut-brain axis hypothesis. We further validated our results using WGS data from the GEMMA cohort, and discovered differences in proportions between autistic probands and familial controls in an independent validation cohort.

Acknowledgements

The work was supported by European Commission Horizon 2020 programme (grant no. 825033)

Conflicts of Interest

The authors have no conflicts of interests.

Author contributions

JL, KS, CL-A, CR, MN and RA designed the study. NV, KS, ST, RA and JL performed the data analysis and integration. KS, NV, ST, RA and JL conducted the validation and figure generation. KS, NV, AF, CR, RA and JL wrote the paper. All authors contributed to critical revisions and approved the final manuscript.

Figure Legends:

Figure 1: Workflow for the analyses conducted in the study. Multivariate GWAS was performed on selected GWAS studies including autism and 11 co-occurring traits: ADHD, ADHD childhood, anxiety, depression, digestion, educational attainment, obsessive compulsive disorder, panic attack, social phobia, IBD and epilepsy. 22 SNPs were selected and evaluated with Colocalization and Mendelian Randomization. Further validation of the selected SNPs utilized independent datasets from Satterstrom et.al (2019) as well as the GEMMA -study. The GEMMA whole genome sequencing (WGS) processing included variant calling to infer structural and single nucleotide variants (SVs and SNVs) present in the samples.

Figure 2: Results from the post GWAS analysis of the 22 selected SNPs. **a,b)** Colocalization processing using the original summary statistics of autism and EA for (a) rs62061734 (MAPT, failed colocalization with H4 probability 8.19%, p-value 0.09), and (b) rs538628 (NSF, passed colocalization with H4 probability 100%, p-value 1.1e-05), depicting supporting regional SNPs (x-axis) and their negative log10 p-value (y-axis) and beta effect direction (circles negative, triangles positive). **c,d)** Mendelian randomization (MR) results using inverse variance weighted (IVW) -method for association of autism SNP effects (y-axis) and EA SNP effects (x-axis) in the NSF (c) and MAPT (d)s -gene. **e)** Pathway analysis for the genes associated with the selected SNPs show enrichment in processes related to neurons using Reactome database. The length of the bar represents the significance of that specific gene-set or pathway and the color indicates the significance of the pathway. Details of the pathways and genes with their associated p-values are listed in Supplementary Table 8. **f)** Organ system enrichment was applied using WebCSEA, using the 22 central trait genes and found significance (p-value < 1e-03) with the digestive, nervous and sensory organ systems.

522 **Figure 3**: Results from the validation with GEMMA -cohort. Heatmap shows the clustering of the
523 selected SNPs in the cohort, with rows showing the genes associated to each SNP and samples
524 shown in columns. Samples with structural variation in chr17q21.31 can be seen to cluster together
525 on the left. Green annotation visualizes the number of breakends (BNDs) detected in the 900KB
526 region, which are suggestive of an inversion. Darker yellow annotation highlights the samples with
527 the 238bp deletion located in intron 9 of MAPT -gene, also associated with the H2 haplotype.
528 Assessment of the H1/H2 haplotype for MAPT was further evaluated using the SNP rs1052553.
529 H1 haplotype has been shown to increase risk of other neurological disorders such as Altzheimer
530 and Parkinson's Disease, and as seen in the GEMMA cohort, autism probands tend to express the
531 H1 haplotype in darker purple (10 out of 13).

532

533 Supplementary Figure Legends:

534

535 Supplementary Figure 1: Genetic correlation of the traits included in the analysis. ADHD=Atten-
536 tion Deficit Hyperactivity Disorder, OCD=Obsessive Compulsive Disorder, IBS=Irritable Bowel
537 Syndrome.
538 Supplementary Figure 2: Heatmap clusters and visualizes the autism co-occurring trait
539 importance, with blue denoting more importance as seen for EA, ADHD and ADHDCHILD.
540 ADHD=Attention Deficit Hyperactivity Disorder, OCD=Obsessive Compulsive Disorder,
541 IBS=Irritable Bowel Syndrome.
542 Supplementary Figure 3: P-value and BIC decomposition processing of MAPT and NSF to iden-
543 tify autism central traits. PanicA=Panic Attack, SocialP=Social Phobia, ADHD=Attention Defi-
544 cit Hyperactivity Disorder, OCD=Obsessive Compulsive Disorder, IBS=Irritable Bowel Syn-
545 drome, EA=Educational Attainment.
546 Supplementary Figure 4: Pathway analysis using the Elsevier database also highlights neuronal
547 processes, with bar length and color the same as in e. Pathways marked with (H) are hypothe-
548 sized, with more details listed in Supplementary Table8.
549 Supplementary Figure 5: Tissue and cell (TS) type enrichment using WebCSEA and the list of
550 the 22 central trait genes found that the most enriched tissues are related to cerebrum, cortex and

551 small intestine related tissue types. Lake 2017 refers to data from human brain single cell analy-
552 sis project (https://pubmed.ncbi.nlm.nih.gov/29227469/) while HCA stands for histologic chori-
553 oamnionitis, an intrauterine inflammatory condition.
554 Supplementary Figure 6: Autism multivariate GWAS associations within the MAPT H1/H2
555 haplotype, 17q21 arm region, are presented in a Manhattan plot, in the context of Grove et al.
556 GWAS results.
557
558 **Supplementary Table Data Titles**
559 Supplementary Table 1: Multivariate associations with autism identified as a central trait as
560 defined by MetaPhat decomposition. As marked with asterisks, 10/22 associations have been
561 reported in GWAS catalog (not ASD) while 3 genes, in bold, are identified as SFARI autism
562 genes
563
564 Supplementary Table 2: Number of LD adjusted SNPs applied within the 22 selected association
565 regions for colocalization and mendelian randomization analyses
566
567 Supplementary Table 3: Colocalization results for the GWAS SNPs
568
569 Supplementary Table 4:(A) Mendelian randomization (MR) results for the selected autism
570 central SNPs. (B) Multivariable Mendelian randomization (MR) results for applicable
571 associations with 3+ central traits
572
573 Supplementary Table 5: 22 Genes found in genesets are curated/implicated with gut microbiome
574 and neural systems from Gene Cards
575
576 Supplementary Table 6: List of 380 Significant SNPs (p-value < 5e-8) identified by MetaPhat
577 multivariate-GWAS using autism and 11 autism co-ocurring condition summary statistics
578
579 Supplementary Table 7: 203/380 multivariate GWAS SNPs are already reported in the GWAS
580 catalog while 10 of the selected 22 (autismv central trait, in bold) SNPs are found
581

582 Supplementary Table 8: Top 25 pathways, based on P-value of Pathway enrichment using
583 Enrichr towards Elsevier (a) and Reactome (b) pathways databases using the 22 genes mapped to
584 the autism central are listed
585
586 Supplementary Table 9: 22 autism central SNPs are validated with 53 GEMMA genotypes called
587 from WGS

588 # References:

589 Aaldijk Emma and Vermeiren Yannick, 2022, "The role of serotonin within the microbiota-gut-
590 brain axis in the development of Alzheimer's disease: A narrative review", Ageing Research
591 Reviews, Volume 75, 2022,101556, ISSN 1568-1637,
592 https://doi.org/10.1016/j.arr.2021.101556

593 Arpi, Magdalena Navarro Torres, and T. Ian Simpson. 2022. "SFARI Genes and Where to Find
594 Them; Modelling Autism Spectrum Disorder Specific Gene Expression Dysregulation with
595 RNA-Seq Data." *Scientific Reports* 12 (1): 10158.
596 Ashburner, Jill, Jenny Ziviani, and Sylvia Rodger. 2010. "Surviving in the Mainstream: Capacity
597 of Children with Autism Spectrum Disorders to Perform Academically and Regulate Their
598 Emotions and Behavior at School." *Research in Autism Spectrum Disorders* 4 (1): 18–27.
599 Avila, Jesus, Jose J. Lucas, Mar Perez, and Felix Hernandez. 2004. "Role of Tau Protein in Both
600 Physiological and Pathological Conditions." *Physiological Reviews* 84 (2): 361–84.
601 Baio, Jon, Lisa Wiggins, Deborah L. Christensen, Matthew J. Maenner, Julie Daniels, Zachary
602 Warren, Margaret Kurzius-Spencer, et al. 2018. "Prevalence of Autism Spectrum Disorder
603 Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring
604 Network, 11 Sites, United States, 2014." *Morbidity and Mortality Weekly Report.*
605 *Surveillance Summaries* 67 (6): 1–23.
606 Bonfiglio, Ferdinando et al. "GWAS of stool frequency provides insights into gastrointestinal
607 motility and irritable bowel syndrome." Cell genomics vol. 1,3 None. 8 Dec. 2021,
608 doi:10.1016/j.xgen.2021.100069
609 Boorstein, Hilary Corinne. 2008. *Gastrointestinal and Feeding Problems in Young Children: A*
610 *Comparison Between Autism Spectrum Disorders and Other Developmental Delays*.

University of Connecticut.

Bowden, Jack, George Davey Smith, and Stephen Burgess. 2015. "Mendelian Randomization with Invalid Instruments: Effect Estimation and Bias Detection through Egger Regression." *International Journal of Epidemiology* 44 (2): 512–25.

Brion, J. P., J. N. Octave, and A. M. Couck. 1994. "Distribution of the Phosphorylated Microtubule-Associated Protein Tau in Developing Cortical Neurons." *Neuroscience* 63 (3): 895–909.

Chaste, Pauline, and Marion Leboyer. 2012. "Autism Risk Factors: Genes, Environment, and Gene-Environment Interactions." *Dialogues in Clinical Neuroscience* 14 (3): 281–92.

Cheng, Lingsha, Haoqian Wu, Zhuo Chen, Haiping Hao, and Xiao Zheng. 2023. "Gut Microbiome at the Crossroad of Genetic Variants and Behavior Disorders." *Gut Microbes* 15 (1): 2201156.

Chen, Xiaoyu, Ole Schulz-Trieglaff, Richard Shaw, Bret Barnes, Felix Schlesinger, Morten Källberg, Anthony J. Cox, Semyon Kruglyak, and Christopher T. Saunders. 2015. "Manta: Rapid Detection of Structural Variants and Indels for Germline and Cancer Sequencing Applications." *Bioinformatics* 32 (8): 1220–22.

Cichonska, Anna, Juho Rousu, Pekka Marttinen, Antti J. Kangas, Pasi Soininen, Terho Lehtimäki, Olli T. Raitakari, et al. 2016. "metaCCA: Summary Statistics-Based Multivariate Meta-Analysis of Genome-Wide Association Studies Using Canonical Correlation Analysis." *Bioinformatics* 32 (13): 1981–89.

Cochran, W. G. 1941. "The Distribution of the Largest of a Set of Estimated Variances as a Fraction of Their Total." *Annals of Eugenics* 11 (1): 47–52.

Corsi, Andrea, Cristina Bombieri, Maria Teresa Valenti, and Maria Grazia Romanelli. 2022. "Tau Isoforms: Gaining Insight into MAPT Alternative Splicing." *International Journal of Molecular Sciences* 23 (23). https://doi.org/10.3390/ijms232315383.

Cross-Disorder Group of the Psychiatric Genomics Consortium. Electronic address: plee0@mgh.harvard.edu, and Cross-Disorder Group of the Psychiatric Genomics Consortium. 2019. "Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders." *Cell* 179 (7): 1469–82.e11.

Dai, Y., Hu, R., Liu, A., Cho, K. S., Manuel, A. M., Li, X., Dong, X., Jia, P., & Zhao, Z. (2022). WebCSEA: web-based cell-type-specific enrichment analysis of genes. Nucleic acids

research, 50(W1), W782–W790. https://doi.org/10.1093/nar/gkac392

Derkinderen, Pascal, Malvyne Rolli-Derkinderen, Guillaume Chapelet, Michel Neunlist, and Wendy Noble. 2021. "Tau in the Gut, Does It Really Matter?" *Journal of Neurochemistry* 158 (2): 94–104.

English, Adam C., Vipin K. Menon, Richard A. Gibbs, Ginger A. Metcalf, and Fritz J. Sedlazeck. 2022. "Truvari: Refined Structural Variant Comparison Preserves Allelic Diversity." *Genome Biology* 23 (1): 1–20.

Fabregat, Antonio, Steven Jupe, Lisa Matthews, Konstantinos Sidiropoulos, Marc Gillespie, Phani Garapati, Robin Haw, et al. 2018. "The Reactome Pathway Knowledgebase." *Nucleic Acids Research* 46 (D1): D649–55.

Feliciano, Pamela, Xueya Zhou, Irina Astrovskaya, Tychele N. Turner, Tianyun Wang, Leo Brueggeman, Rebecca Barnard, et al. 2019. "Exome Sequencing of 457 Autism Families Recruited Online Provides Evidence for Autism Risk Genes." *NPJ Genomic Medicine* 4 (August): 19.

Fiorentino Maria, Sapone A, Senger S, et al. Blood-brain barrier and intestinal epithelial barrier alterations in autism spectrum disorders. *Mol Autism*. 2016;7:49. Published 2016 Nov 29. doi:10.1186/s13229-016-0110-z

Gaugler, Trent, Lambertus Klei, Stephan J. Sanders, Corneliu A. Bodea, Arthur P. Goldberg, Ann B. Lee, Milind Mahajan, et al. 2014. "Most Genetic Risk for Autism Resides with Common Variation." *Nature Genetics* 46 (8): 881–85.

Giambartolomei, Claudia, Damjan Vukcevic, Eric E. Schadt, Lude Franke, Aroon D. Hingorani, Chris Wallace, and Vincent Plagnol. 2014. "Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics." *PLoS Genetics* 10 (5): e1004383.

Grove, Jakob, Stephan Ripke, Thomas D. Als, Manuel Mattheisen, Raymond K. Walters, Hyejung Won, Jonatan Pallesen, et al. 2019. "Identification of Common Genetic Risk Variants for Autism Spectrum Disorder." *Nature Genetics* 51 (3): 431–44.

Gutowska-Ding, Maria Weronika, Zandra C. Deans, Christophe Roos, Jukka Matilainen, Farrah Khawaja, Kim Brügger, Jo Wook Ahn, Christopher Boustred, and Simon J. Patton. 2020. "One Byte at a Time: Evidencing the Quality of Clinical Service next-Generation Sequencing for Germline and Somatic Variants." *European Journal of Human Genetics:*

673 *EJHG* 28 (2): 202–12.
674 Haycock, Philip C., Stephen Burgess, Kaitlin H. Wade, Jack Bowden, Caroline Relton, and
675 George Davey Smith. 2016. "Best (but Oft-Forgotten) Practices: The Design, Analysis, and
676 Interpretation of Mendelian Randomization Studies." *The American Journal of Clinical*
677 *Nutrition* 103 (4): 965–78.
678 Holingue, Calliope, Carol Newill, Li-Ching Lee, Pankaj J. Pasricha, and M. Daniele Fallin. 2018.
679 "Gastrointestinal Symptoms in Autism Spectrum Disorder: A Review of the Literature on
680 Ascertainment and Prevalence." *Autism Research: Official Journal of the International*
681 *Society for Autism Research* 11 (1): 24–36.
682 Iwata, Keiko, Hideo Matsuzaki, Taro Tachibana, Koji Ohno, Saori Yoshimura, Hironori Taka-
683 mura, Kohei Yamada, et al. 2014. "N-Ethylmaleimide-Sensitive Factor Interacts with the
684 Serotonin Transporter and Modulates Its Trafficking: Implications for Pathophysiology in
685 Autism." *Molecular Autism* 5 (May): 33.
686 Jansen, Arija G., Gwen C. Dieleman, Philip R. Jansen, Frank C. Verhulst, Danielle Posthuma,
687 and Tinca J. C. Polderman. 2020. "Psychiatric Polygenic Risk Scores as Predictor for
688 Attention Deficit/Hyperactivity Disorder and Autism Spectrum Disorder in a Clinical Child
689 and Adolescent Sample." *Behavior Genetics* 50 (4): 203–12.
690 Kern, Janet K., David A. Geier, Lisa K. Sykes, Mark R. Geier, and Richard C. Deth. 2015. "Are
691 ASD and ADHD a Continuum? A Comparison of Pathophysiological Similarities Between
692 the Disorders." *Journal of Attention Disorders* 19 (9): 805–27.
693 Khachadourian, Vahe, Behrang Mahjani, Sven Sandin, Alexander Kolevzon, Joseph D.
694 Buxbaum, Abraham Reichenberg, and Magdalena Janecka. 2023. "Comorbidities in Autism
695 Spectrum Disorder and Their Etiologies." *Translational Psychiatry* 13 (1): 71.
696 Kuleshov, Maxim V., Matthew R. Jones, Andrew D. Rouillard, Nicolas F. Fernandez, Qiaonan
697 Duan, Zichen Wang, Simon Koplev, et al. 2016. "Enrichr: A Comprehensive Gene Set En-
698 richment Analysis Web Server 2016 Update." *Nucleic Acids Research* 44 (W1): W90–97.
699 Lake, B. B., Chen, S., Sos, B. C., Fan, J., Kaeser, G. E., Yung, Y. C., Duong, T. E., Gao, D.,
700 Chun, J., Kharchenko, P. V., & Zhang, K. (2018). Integrative single-cell analysis of
701 transcriptional and epigenetic states in the human adult brain. Nature biotechnology, 36(1),
702 70–80. https://doi.org/10.1038/nbt.4038
703 Li, Heng. 2013. "Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-

MEM.” http://arxiv.org/abs/1303.3997.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. “The Sequence Alignment/Map Format and SAMtools.” *Bioinformatics* 25 (16): 2078.

Lin, Jake, Nina Mars, Yu Fu, Pietari Ripatti, Tuomo Kiiskinen, Taru Tukiainen, Samuli Ripatti, and Matti Pirinen. 2023. “Integration of Biomarker Polygenic Risk Score Improves Prediction of Coronary Heart Disease.” *JACC: Basic to Translational Science*, October. https://doi.org/10.1016/j.jacbts.2023.07.006.

Lin, Jake, Rubina Tabassum, Samuli Ripatti, and Matti Pirinen. 2020. “MetaPhat: Detecting and Decomposing Multivariate Associations From Univariate Genome-Wide Association Statistics.” *Frontiers in Genetics* 11 (May): 431.

Morton, James T., Dong-Min Jin, Robert H. Mills, Yan Shao, Gibraan Rahman, Daniel McDonald, Qiyun Zhu, et al. 2023. “Multi-Level Analysis of the Gut-Brain Axis Shows Autism Spectrum Disorder-Associated Molecular and Microbial Profiles.” *Nature Neuroscience* 26 (7): 1208–17.

Nalls, Mike A., Nathan Pankratz, Christina M. Lill, Chuong B. Do, Dena G. Hernandez, Mohamad Saad, Anita L. DeStefano, et al. 2014. “Large-Scale Meta-Analysis of Genome-Wide Association Data Identifies Six New Risk Loci for Parkinson’s Disease.” *Nature Genetics* 46 (9): 989–93.

Nayar, Kritika, Julia M. Sealock, Nell Maltman, Lauren Bush, Edwin H. Cook, Lea K. Davis, and Molly Losh. 2021. “Elevated Polygenic Burden for Autism Spectrum Disorder Is Associated With the Broad Autism Phenotype in Mothers of Individuals With Autism Spectrum Disorder.” *Biological Psychiatry* 89 (5): 476–85.

Nesterova, Anastasia P., Anton Yuryev, Eugene A. Klimov, Maria Zharkova, Maria Shkrob, Natalia V. Ivanikova, Sergey Sozin, and Vladimir Sobolev. 2019. *Disease Pathways: An Atlas of Human Disease Signaling Pathways*. Elsevier.

Okbay, Aysu, Yeda Wu, Nancy Wang, Hariharan Jayashankar, Michael Bennett, Seyed Moeen Nehzati, Julia Sidorenko, et al. 2022. “Polygenic Prediction of Educational Attainment within and between Families from Genome-Wide Association Analyses in 3 Million Individuals.” *Nature Genetics* 54 (4): 437–49.

735 Pedersen, C. B., J. Bybjerg-Grauholm, M. G. Pedersen, J. Grove, E. Agerbo, M. Bækvad-
736 Hansen, J. B. Poulsen, et al. 2018. "The iPSYCH2012 Case-Cohort Sample: New Directions
737 for Unravelling Genetic and Environmental Architectures of Severe Mental Disorders."
738 *Molecular Psychiatry* 23 (1): 6–14.
739 Peyre, Hugo, Tabea Schoeler, Chaoyu Liu, Camille Michèle Williams, Nicolas Hoertel,
740 Alexandra Havdahl, and Jean-Baptiste Pingault. 2021. "Combining Multivariate Genomic
741 Approaches to Elucidate the Comorbidity between Autism Spectrum Disorder and Attention
742 Deficit Hyperactivity Disorder." *Journal of Child Psychology and Psychiatry, and Allied*
743 *Disciplines* 62 (11): 1285–96.
744 Phillips, A. N., and G. D. Smith. 1991. "How Independent Are 'Independent' Effects? Relative
745 Risk Estimation When Correlated Exposures Are Measured Imprecisely." *Journal of*
746 *Clinical Epidemiology* 44 (11): 1223–31.
747 Pinto, Dalila, Elsa Delaby, Daniele Merico, Mafalda Barbosa, Alison Merikangas, Lambertus
748 Klei, Bhooma Thiruvahindrapuram, et al. 2014. "Convergence of Genes and Cellular
749 Pathways Dysregulated in Autism Spectrum Disorders." *American Journal of Human*
750 *Genetics* 94 (5): 677–94.
751 Piovesan, Allison, Maria Caracausi, Francesca Antonaros, Maria Chiara Pelleri, and Lorenza
752 Vitale. 2016. "GeneBase 1.1: A Tool to Summarize Data from NCBI Gene Datasets and Its
753 Application to an Update of Human Gene Statistics." *Database: The Journal of Biological*
754 *Databases and Curation* 2016 (December). https://doi.org/10.1093/database/baw153.
755 Poplin, Ryan, Pi-Chuan Chang, David Alexander, Scott Schwartz, Thomas Colthurst, Alexander
756 Ku, Dan Newburger, et al. 2018. "A Universal SNP and Small-Indel Variant Caller Using
757 Deep Neural Networks." *Nature Biotechnology* 36 (10): 983–87.
758 Rees, Jessica M. B., Angela M. Wood, and Stephen Burgess. 2017. "Extending the MR-Egger
759 Method for Multivariable Mendelian Randomization to Correct for Both Measured and
760 Unmeasured Pleiotropy." *Statistics in Medicine* 36 (29): 4705–18.
761 Risch, N., and K. Merikangas. 1996. "The Future of Genetic Studies of Complex Human
762 Diseases." *Science* 273 (5281): 1516–17.
763 Rojas-Charry, Liliana, Leonardo Nardi, Axel Methner, and Michael J. Schmeisser. 2021.
764 "Abnormalities of Synaptic Mitochondria in Autism Spectrum Disorder and Related
765 Neurodevelopmental Disorders." *Journal of Molecular Medicine* 99 (2): 161–78.

766 Romero, Marina, Juan Manuel Aguilar, Ángel Del-Rey-Mejías, Fermín Mayoral, Marta Rapado,
767 Marta Peciña, Miguel Ángel Barbancho, Miguel Ruiz-Veguilla, and José Pablo Lara. 2016.
768 "Psychiatric Comorbidities in Autism Spectrum Disorder: A Comparative Study between
769 DSM-IV-TR and DSM-5 Diagnosis." *International Journal of Clinical and Health*
770 *Psychology: IJCHP* 16 (3): 266–75.
771 Ruotsalainen, Sanni E., Juulia J. Partanen, Anna Cichonska, Jake Lin, Christian Benner, Ida
772 Surakka, FinnGen, et al. 2021. "An Expanded Analysis Framework for Multivariate GWAS
773 Connects Inflammatory Biomarkers to Functional Variants and Disease." *European Journal*
774 *of Human Genetics: EJHG* 29 (2): 309–24.
775 Safran, Marilyn, Irina Dalah, Justin Alexander, Naomi Rosen, Tsippi Iny Stein, Michael
776 Shmoish, Noam Nativ, et al. 2010. "GeneCards Version 3: The Human Gene Integrator."
777 *Database: The Journal of Biological Databases and Curation* 2010 (August): baq020.
778 Satterstrom, F. Kyle, Jack A. Kosmicki, Jiebiao Wang, Michael S. Breen, Silvia De Rubeis,
779 Joon-Yong An, Minshi Peng, et al. 2020. "Large-Scale Exome Sequencing Study Implicates
780 Both Developmental and Functional Changes in the Neurobiology of Autism." *Cell* 180 (3):
781 568–84.e23.
782 Savage, Jeanne E., Philip R. Jansen, Sven Stringer, Kyoko Watanabe, Julien Bryois, Christiaan
783 A. de Leeuw, Mats Nagel, et al. 2018. "Genome-Wide Association Meta-Analysis in
784 269,867 Individuals Identifies New Genetic and Functional Links to Intelligence." *Nature*
785 *Genetics* 50 (7): 912–19.
786 Schaaf, Christian P., Catalina Betancur, Ryan K. C. Yuen, Jeremy R. Parr, David H. Skuse,
787 Louise Gallagher, Raphael A. Bernier, et al. 2020. "A Framework for an Evidence-Based
788 Gene List Relevant to Autism Spectrum Disorder." *Nature Reviews. Genetics* 21 (6): 367–
789 76.
790 Schwarz, Gideon. 1978. "Estimating the Dimension of a Model." *Annals of Statistics* 6 (2): 461–
791 64.
792 Singh, Katyayani, Mohan Jayaram, Maria Kaare, Este Leidmaa, Toomas Jagomäe, Indrek
793 Heinla, Miriam A. Hickey, et al. 2019. "Neural Cell Adhesion Molecule Negr1 Deficiency
794 in Mouse Results in Structural Brain Endophenotypes and Behavioral Deviations Related to
795 Psychiatric Disorders." *Scientific Reports* 9 (1): 5457.
796 Sullivan, Patrick F., Arpana Agrawal, Cynthia M. Bulik, Ole A. Andreassen, Anders D.

Børglum, Gerome Breen, Sven Cichon, et al. 2018. “Psychiatric Genomics: An Update and an Agenda.” *The American Journal of Psychiatry* 175 (1): 15–27.

Suzuki, Hisato, Takeshi Yoshida, Naoya Morisada, Tomoko Uehara, Kenjiro Kosaki, Katsunori Sato, Kohei Matsubara, Toshiyuki Takano-Shimizu, and Toshiki Takenouchi. 2019. “De Novo NSF Mutations Cause Early Infantile Epileptic Encephalopathy.” *Annals of Clinical and Translational Neurology* 6 (11): 2334–39.

Tamlander, Max, Nina Mars, Matti Pirinen, FinnGen, Elisabeth Widén, and Samuli Ripatti. 2022. “Integration of Questionnaire-Based Risk Factors Improves Polygenic Risk Scores for Human Coronary Heart Disease and Type 2 Diabetes.” *Communications Biology* 5 (1): 158.

Tamm, Leanne, Amie Duncan, Aaron Vaughn, Rhyanne McDade, Nicole Estell, Allison Birnschein, and Lori Crosby. 2020. “Academic Needs in Middle School: Perspectives of Parents and Youth with Autism.” *Journal of Autism and Developmental Disorders* 50 (9): 3126–39.

Trampush, J. W., M. L. Z. Yang, J. Yu, E. Knowles, G. Davies, D. C. Liewald, J. M. Starr, et al. 2017. “GWAS Meta-Analysis Reveals Novel Loci and Genetic Correlates for General Cognitive Function: A Report from the COGENT Consortium.” *Molecular Psychiatry* 22 (11): 1651–52.

Troisi, Jacopo, Reija Autio, Thanos Beopoulos, Carmela Bravaccio, Federica Carraturo, Giulio Corrivetti, Stephen Cunningham, et al. 2020. “Genome, Environment, Microbiome and Metabolome in Autism (GEMMA) Study Design: Biomarkers Identification for Precision Treatment and Primary Prevention of Autism Spectrum Disorders by an Integrated Multi-Omics Systems Biology Approach.” *Brain Sciences* 10 (10). https://doi.org/10.3390/brainsci10100743.

Wallace, Chris. 2020. “Eliciting Priors and Relaxing the Single Causal Variant Assumption in Colocalisation Analyses.” *PLoS Genetics* 16 (4): e1008720.

———. 2021. “A More Accurate Method for Colocalisation Analysis Allowing for Multiple Causal Variants.” *PLoS Genetics* 17 (9): e1009440.

Wang, Yipeng, and Eckhard Mandelkow. 2016. “Tau in Physiology and Pathology.” *Nature Reviews. Neuroscience* 17 (1): 5–21.

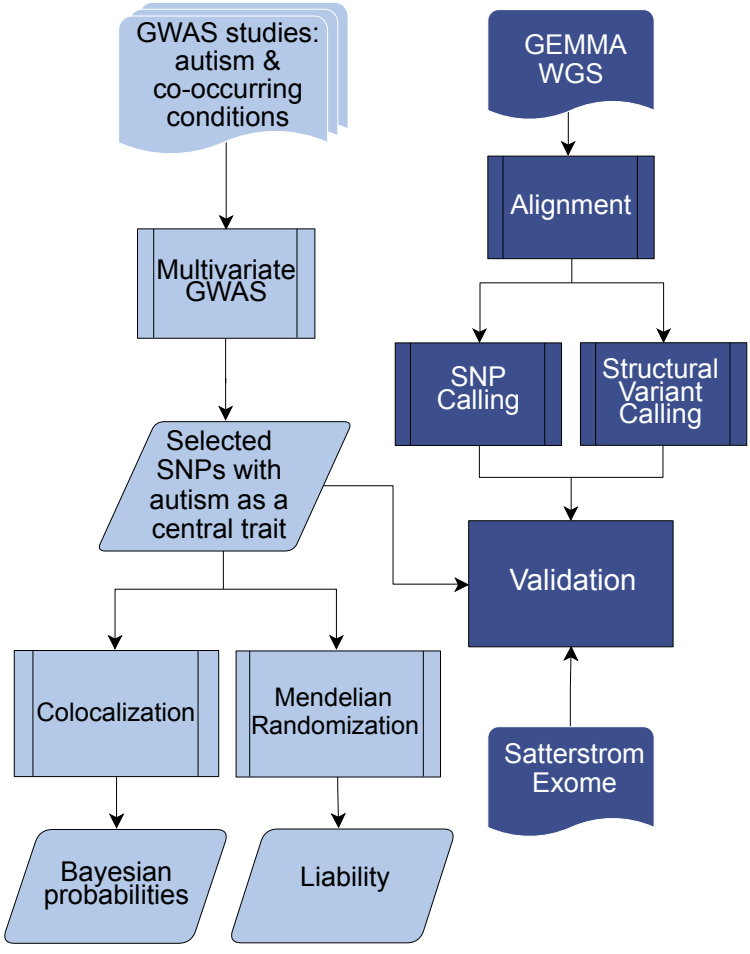
Wider, C., C. Vilariño-Güell, B. Jasinska-Myga, M. G. Heckman, A. I. Soto-Ortolaza, S. A. Cobb, J. O. Aasly, et al. 2010. “Association of the MAPT Locus with Parkinson’s Disease.”

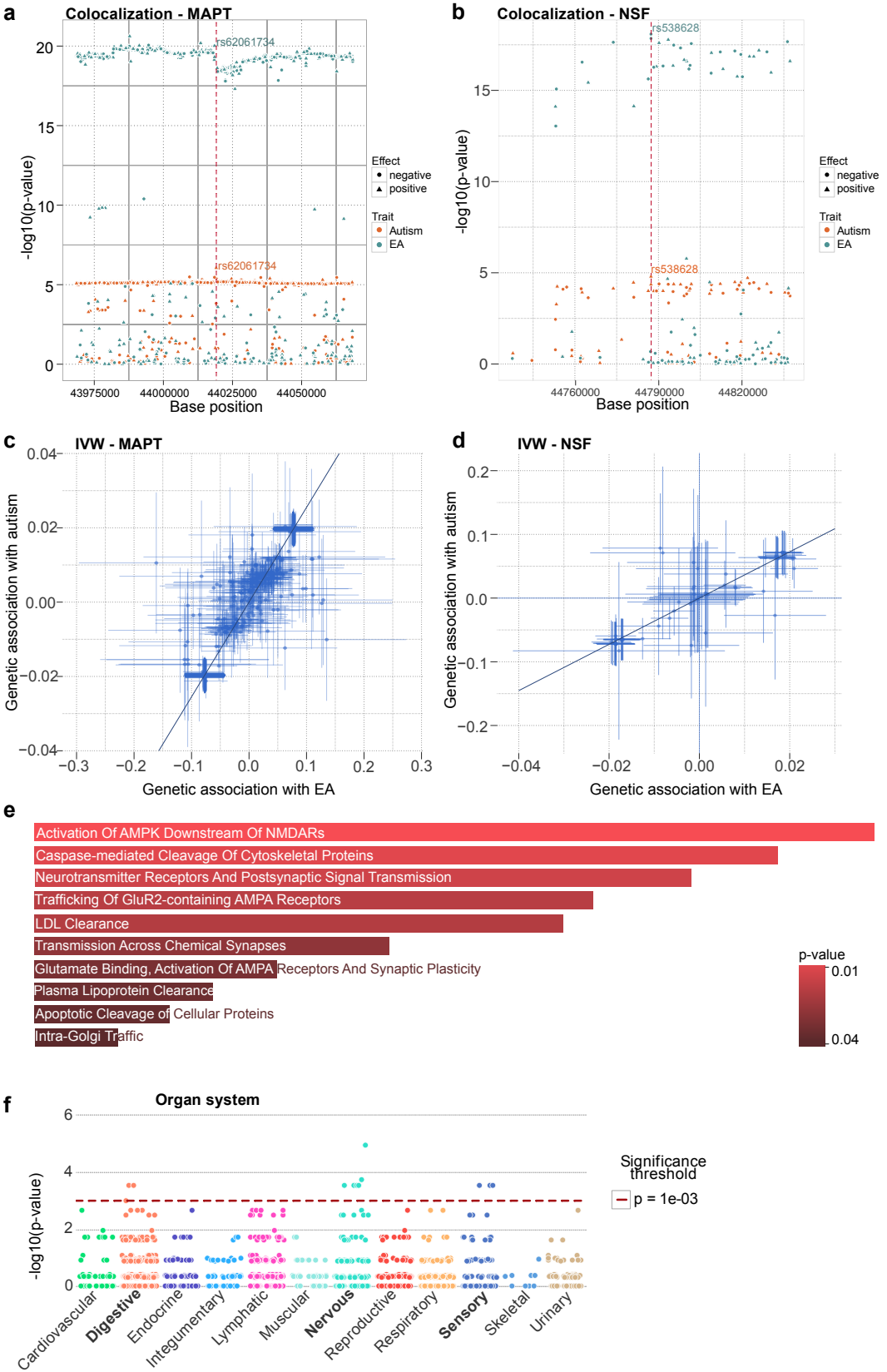
European Journal of Neurology: The Official Journal of the European Federation of Neurological Societies 17 (3): 483–86.

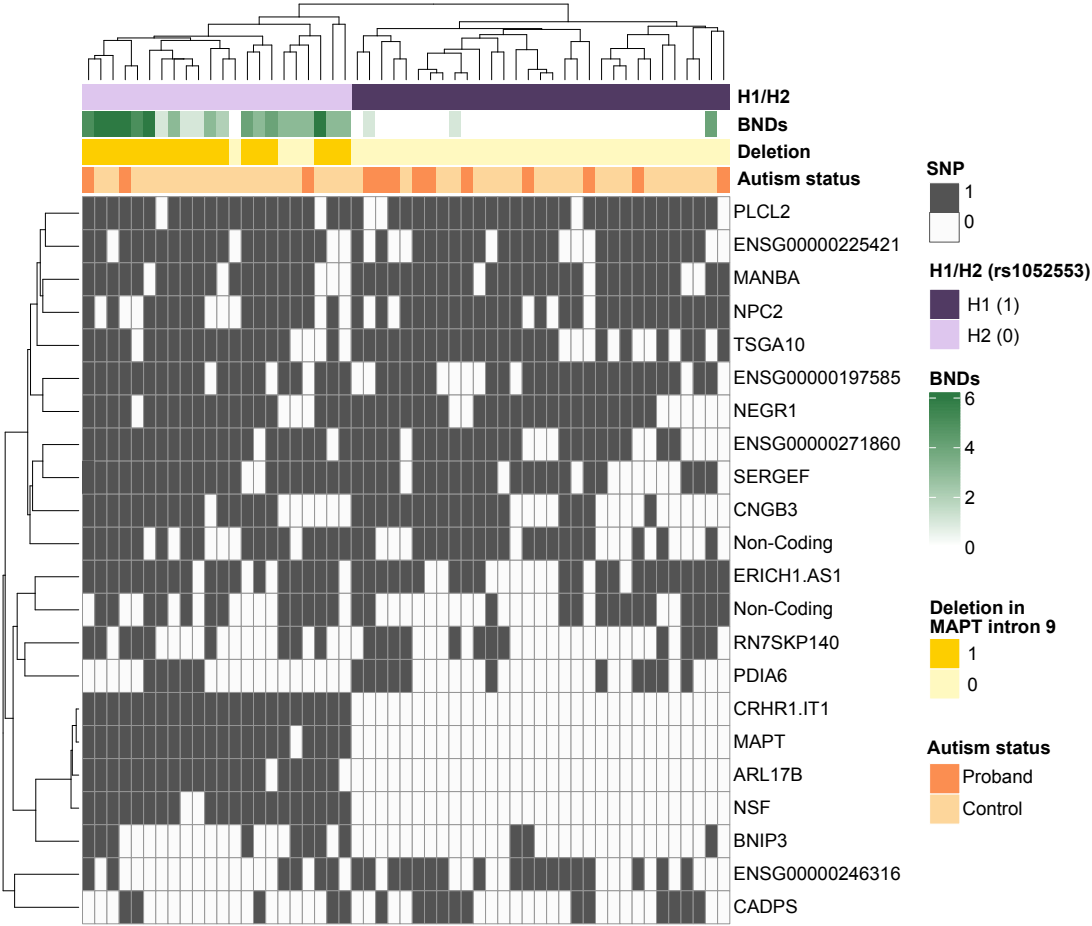
Xie, Min-Jue, Keiko Iwata, Yasuyuki Ishikawa, Yuki Nomura, Tomomi Tani, Koshi Murata, Yugo Fukazawa, and Hideo Matsuzaki. 2021. "Autistic-Like Behavior and Impairment of Serotonin Transporter and AMPA Receptor Trafficking in N-Ethylmaleimide Sensitive Factor Gene-Deficient Mice." *Frontiers in Genetics* 12 (October): 748627.

Yavorska, Olena O., and Stephen Burgess. 2017. "MendelianRandomization: An R Package for Performing Mendelian Randomization Analyses Using Summarized Data." *International Journal of Epidemiology* 46 (6): 1734–39.

Yun, Taedong, Helen Li, Pi-Chuan Chang, Michael F. Lin, Andrew Carroll, and Cory Y. McLean. 2021. "Accurate, Scalable Cohort Variant Calls Using DeepVariant and GLnexus." *Bioinformatics* 36 (24): 5582–89.


GWAS studies: autism & co-occurring conditions
Multivariate GWAS
Selected SNPs with autism as a central trait
Colocalization
Mendelian Randomization
Bayesian probabilities
Liability
GEMMA WGS
Alignment
SNP Calling
Structural Variant Calling
Validation
Satterstrom Exome


a
Colocalization - MAPT
rs62061734
rs62061734
-log10(p-value)
Base position
Effect
negative
positive
Trait
Autism
EA
b
Colocalization - NSF
rs538628
rs538628
-log10(p-value)
Base position
Effect
negative
positive
Trait
Autism
EA
c
IVW - MAPT
Genetic association with autism
Genetic association with EA
d
IVW - NSF
Genetic association with autism
Genetic association with EA
e
Activation Of AMPK Downstream Of NMDARs
Caspase-mediated Cleavage Of Cytoskeletal Proteins
Neurotransmitter Receptors And Postsynaptic Signal Transmission
Trafficking Of GluR2-containing AMPA Receptors
LDL Clearance
Transmission Across Chemical Synapses
Glutamate Binding, Activation Of AMPA Receptors And Synaptic Plasticity
Plasma Lipoprotein Clearance
Apoptotic Cleavage of Cellular Proteins
Intra-Golgi Traffic
p-value
0.01
0.04
f
Organ system
-log10(p-value)
Significance threshold
p = 1e-03
Cardiovascular
Digestive
Endocrine
Integumentary
Lymphatic
Muscular
Nervous
Reproductive
Respiratory
Sensory
Skeletal
Urinary


H1/H2
BNDs
Deletion
Autism status
PLCL2
ENSG00000225421
MANBA
NPC2
TSGA10
ENSG00000197585
NEGR1
ENSG00000271860
SERGEF
CNGB3
Non-Coding
ERICH1.AS1
Non-Coding
RN7SKP140
PDIA6
CRHR1.IT1
MAPT
ARL17B
NSF
BNIP3
ENSG00000246316
CADPS
SNP
1
0
H1/H2 (rs1052553)
H1 (1)
H2 (0)
BNDs
6
4
2
0
Deletion in MAPT intron 9
1
0
Autism status
Proband
Control