

RESEARCH

Open Access



Exploring autism spectrum disorder and co-occurring trait associations to elucidate multivariate genetic mechanisms and insights

Karoliina Salenius^{1†}, Niina Väljä^{1†}, Sini Thusberg¹, Francois Iris³, Christine Ladd-Acosta⁴, Christophe Roos⁵, Matti Nykter^{1,6}, Alessio Fasano^{7,8}, Reija Autio^{9†}, Jake Lin^{1,10*†} on behalf of the GEMMA study²

Abstract

Background Autism spectrum disorder (ASD) is a partially heritable neurodevelopmental trait, and people with ASD may also have other co-occurring trait such as ADHD, anxiety disorders, depression, mental health issues, learning difficulty, physical health traits and communication challenges. The concomitant development of ASD and other neurological traits is assumed to result from a complex interplay between genetics and the environment. However, only a limited number of studies have performed multivariate genome-wide association studies (GWAS) for ASD.

Methods We conducted to-date the largest multivariate GWAS on ASD and 8 ASD co-occurring traits (ADHD, ADHD childhood, anxiety stress (ASDR), bipolar (BIP), disruptive behaviour (DBD), educational attainment (EA), major depression, and schizophrenia (SCZ)) using summary statistics from leading studies. 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 ASD. To further validate our findings, pathway and quantified trait loci (QTL) resources as well as independent datasets consisting of 112 (45 probands) whole genome sequence data from the GEMMA project were utilized.

Results Multivariate GWAS resulted in 637 significant associations ($p < 5e-8$), among which 322 are reported for the first time for any trait. 37 SNPs were identified to contain ASD and one or more traits in their central trait set, including variants mapped to known SFARI ASD genes *MAPT*, *CADPS* and *NEGR1* as well as novel ASD genes *KANSL1*, *NSF* and *NTM*, associated with immune response, synaptic transmission, and neurite growth respectively. Mendelian randomization analyses found that genetic liability for ADHD childhood, ASRD and DBT has causal effects on the risk of ASD while genetic liability for ASD has causal effects on the risk of ADHD, ADHD childhood, BIP, WA, MDD and SCZ. Frequency differences of SNPs found in *NTM* and *CADPS* genes, respectively associated with neurite growth and neural/endocrine calcium regulation, were found between GEMMA ASD probands and controls. Pathway, QTL and cell type enrichment implicated microbiome, enteric inflammation, and central nervous system enrichments.

Conclusions Our study, combining multivariate GWAS with systematic decomposition, identified novel genetic associations related to ASD and ASD co-occurring driver traits. Statistical tests were applied to discern evidence for shared and interpretable liability between ASD and co-occurring traits. These findings expand upon the current

[†]Karoliina Salenius, Niina Väljä, Reija Autio and Jake Lin contributed equally to this work.

*Correspondence:

Jake Lin

jake.lin@tuni.fi

Full list of author information is available at the end of the article



understanding of the complex genetics regulating ASD and reveal insights of neuronal brain disruptions potentially driving development and manifestation.

Highlights

Multivariate GWAS resulted in 637 significant ASD associations ($p < 5e-8$), among which 322 are reported for the first time.

The novel associations mapped to known SFARI ASD genes *CADPS*, *MAPT* and *NEGR1* and novel ASD genes *KANSL1*, *NSF* and *NTM*, associated with immune response, synaptic transmission, and neurite growth, potentially driving the gut brain-barrier hypothesis underpinning ASD development.

CuONPs induce co-occurrence of autophagy activation and autophagic flux blockade.

Mendelian randomization analyses found that genetic liability for ASRD and DBT have causal effects on the risk of ASD while genetic liability for ASD have causal effects on the risk of ADHD, BIP, WA, MDD and SCZ. Bidirectional genetic liability causal effects were confirmed between ASD and ADHD childhood.

Keywords ASD, ASD genetically correlated traits, Multivariate GWAS, Mendelian randomization, GEMMA

Introduction

ASD spectrum disorders (ASD) is an umbrella term for a group of heterogeneous neurodevelopmental traits that manifest in early childhood. ASD is a complex disorder with both genetic and environmental risk factors [10, 30, 45]. The diagnosis of ASD is based on 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 traits in autistic persons are attention deficit hyperactivity disorder (ADHD), ADHD childhood, anxiety, bipolar (BP), depression, epilepsy, obsessive compulsive disorders (OCD) and stress related traits, all of which share overlapping diagnostic attributes and challenging symptoms with ASD [30, 57]. 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) [4, 68]. Some children with ASD having higher IQ scores also comparatively experience harder academic struggles due to co-occurring traits and difficulties in social interactions [3].

Together with recent advances in genomics technology and pivotal support from the engaged ASD community, 1,162 genes are currently implicated with ASD development and these are curated in the SFARI [2, 19, 52] gene module. These genes, with varying degrees of effect, are scored using the Evaluation of ASD Gene Link Evidence (EAGLE) framework [61]. Surprisingly, while it is known that common variants contribute to most of the genetic background [18], 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

ASD cases and 27,969 controls, where 12 significant variant associations were reported [19].

Given that there is overlap in symptoms between ASD and ADHD, recent genetics studies found shared genetic factors underlying ASD and ADHD [40, 41, 50], with partial concordance between bidirectional colocalization single nucleotide variants (SNPs). However, these studies were 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 [5]. Recently, there have been promising results that link microbiome disruption and diversity [44] as a novel contributing factor to ASD. While Grove and colleagues found that 7 of the 12 ASD SNP associations have similar significance towards EA and psychosis traits depression and schizophrenia [19], still little is known concerning the joint liability and the shared genetic mechanisms between ASD and ASD co-occurring traits including ADHD, ADHD childhood, anxiety-stress related disorder (ASRD), bipolar, disruptive behavior disorder (DBD), EA, epilepsy, inflammatory bowel disease (IBD), major depression, obsessive compulsive disorder (OCD) and schizophrenia (SCZ). Respectively, the 11 co-occurring trait summary statistics are retrieved from large reputable cohorts, listed in Table 1 and Supplementary Table 1.

To attenuate the genetic knowledge gaps in ASD and expand the exploration of potential shared co-occurring trait genetic associations, this study performed multivariate genome-wide association study (GWAS) with summary statistics from ASD and 11 co-occurring traits from large reputable cohorts. To achieve this, colocalization (coloc) was systematically applied to test the robustness between the shared variants and traits [75]. Mendelian randomisation (MR) was further applied, using the

multivariate variants and the essential traits, to assess liability relationships between ASD and the selected co-occurring traits [6, 55]. This study seeks to further clarify functional, regulatory and tissue type differentiation with enrichment and integration of quantified trait loci (QTL) while validating our key findings with independently sequenced genomes from the GEMMA cohort [70].

Methods and materials

GWAS summary statistics for ASD and ADHD were collected from the Psychiatric Genomics Consortium (PGC) and iPSYCH [49, 65] studies. Education attainment [47] summary file was collected from the Social Science Genetic Association Consortium (SSGAC). Additional ASD co-occurring traits, selected based on LDSC (LD Score Regression) genetic correlation

(p -value < 0.05) with ASD, include ADHD childhood, bipolar (BP), anxiety-stress disorder (ASRD), disruptive behaviour (DBD), major depression (MDD) and schizophrenia (SCZ), with sample sizes ranging from 31,890–765,283 are shown in Table 1 (additional details including doi references listed Supplementary Table 1). To estimate potential sample overlaps, pairwise LDSC intercepts with ASD are calculated and reported in Supplementary Table 1. Summary statistics are joined, yielding 4,525,476 SNPs, and applied in a multivariate GWAS setting. Follow-up analysis includes decomposition aiming to detect the most important traits while colocalization and Mendelian randomisation analysis are conducted to explore shared liability as shown in Fig. 1.

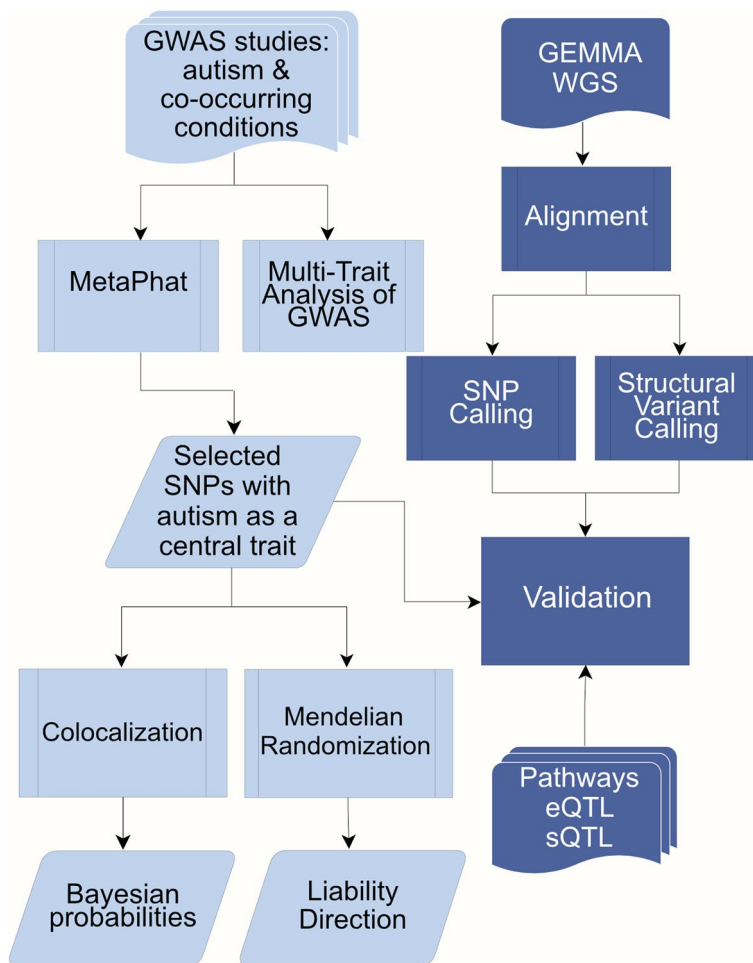


Fig. 1 Workflow for the analyses conducted in the study. Multivariate GWAS was performed on selected GWAS studies including ASD and 8 co-occurring traits: ADHD, ADHD childhood, bipolar, anxiety, disruptive behaviour, educational attainment, major depression and schizophrenia. 37 SNPs were selected and evaluated with Colocalization and Mendelian Randomization. Further validation of these SNPs utilized pathway and EBI eQTL/sQTL catalogs 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

Table 1 Data of ASD and 8 genetically correlated traits ($P < 0.05$, calculated from LDSC), reported SNP heritability (H2), genetic covariance and covariance scores (standard errors) are presented and applied towards multivariate-GWAS to explore multivariate associations and additional trait refinement. More details and excluded traits are listed in supplementary table 1

Trait	Heritability (H2)	Genetic correlation (rg)	P	Genetic covariance	Intercept
ASD ^a	0.118 (0.010)	na	na	na	na
ADHD ^b	0.140 (0.010)	0.535 (0.041)	1.44e-38	0.074(0.006)	0.233 (0.009)
ADHDCHILD ^c	0.235 (0.015)	0.478 (0.052)	5.21e-20	0.104 (0.012)	0.260 (0.007)
ASRD ^d	0.280 (0.027)	0.441 (0.079)	2.22e-08	0.090 (0.015)	0.221 (0.006)
Bipolar ^e	0.068 (0.003)	0.219 (0.041)	9.67e-08	0.026 (0.005)	0.032 (0.006)
DBD ^f	0.100 (0.012)	0.186 (0.07)	0.008	0.026 (0.010)	0.179 (0.006)
EA ^g	0.321 (0.009)	0.207 (0.025)	9.95e-17	0.053 (0.007)	-0.005 (0.007)
MDD ^h	0.090 (0.004)	0.505 (0.003)	2.78e-36	0.037 (0.003)	0.155 (0.005)
SCZ ⁱ	0.240 (0.007)	0.258 (0.035)	7.87e-14	0.070 (0.010)	0.018 (0.007)

Abbreviations: ADHD Attention Deficit Hyper Disorder, ADHDCHILD ADHD childhood, ASRD Anxiety-Stress Disorder, DBD Disruptive Behaviour Disorder, EA Education attainment, MDD Major Depressive Disorder, SCZ Schizophrenia

^a Grove et al. [19] (PMID: 30804558)

^b iPSYCH + deCODE + PGC, Demontis et al. 2023 (PMID: 36702997)

^c iPSYCH, Rajagopal et al. 2022 (PMID: 35927488)

^d iPSYCH excluding ASD cases Meier et al. 2019 (PMID: PMC6537792)

^e Discovery excluding UKB Mullins et al. 2021 (PMID: 34002096)

^f Demontis et al. 2021 (PMID: 33495439)

^g Discovery cohorts excluding 23andme, Okbay et al. [47] (PMID: 35361970)

^h PGC excluding UKB, Wray et al. 2018 (PMID: 29700475)

ⁱ PGC Wave 3, Trutbetkoy et al. 2022 (PMID: 35396580)

Multivariate GWAS and determination of central traits

Multivariate GWAS on ASD and ASD co-occurring 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 [12, 36, 58]. The objective of metaCCA is to find the optimal genetic effect combination that is maximally correlated with a linear combination of the trait variables. ASD multivariate central traits are identified by MetaPhat decomposition based on iterative tracing of p -values (p) from trait subsets (relative to $5e-8$) and Bayesian Information Criterion (BIC) [62] representing model fit. Essentially, driver trait(s) are the subsets of the multivariate association that drives the p -value, and without the drivers, the multivariate association is no longer significant ($p > 5e-8$). Similarly, as the decomposition processing is exhaustive (iterates from k to 1), an optimal subset is identified by comparing BIC values [36]. For simplicity, the central traits are the union of the driver and optimal BIC traits. Multi-Trait Analysis of GWAS (MTAG) [71], a high performance multivariate-GWAS that addresses sample overlap, is additionally performed for validation.

Genetic annotations, pathway enrichment and validation

SFARI Base Gene resource, GeneCards and GWAS catalog were used to assess the novelty of variants and genes

associated with ASD [2, 39, 60]. snpXplorer was applied towards SNP annotation [69]. Reactome and WikiPathway databases pathway enrichments were evaluated with the Enrichr tool [31]. Human organ and cell type systems enrichment analysis, encompassing 1,466 tissue-cell type and single-cell RNAseq panels, was conducted using WebCSEA [13, 33]. eQTL and sQTL were assessed within the QTL catalog, via FIVEx portal [32].

Colocalization analyses

Colocalization was performed for the selected multivariate ASD SNPs to assess if the associated variants in the locus are shared genetically between ASD and the 8 co-occurring related ASD 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). The analyses were performed using the R package coloc [20, 33].

The colocalization analysis was conducted using the absolute base factor colocalization method (coloc.abf), which is a Bayesian colocalization analysis method. A region size window of 100KB (± 50 KB flanking the SNP position) was selected to comprehensively span potential LD and regulatory elements [53]. The different hypotheses tested include: H0 (no liable variant), H1 (liable variant only for trait 1), H2 (liable variant only for trait 2), H3 (two separate liable variants), H4 (common liable variant

shared between the traits). As recommended [74], default setting prior probability thresholds were applied: $1e-4$ for H1, H2 and H3 and $1e-5$ for H4 while posterior probability ($H4 > 90\%$) is conservatively applied to estimate shared liability.

Mendelian randomization analyses

Mendelian Randomization analyses (MR) was conducted on the selected multivariate GWAS SNPs based on their assigned central traits, to explore the liability, direction and independent (reverse causation) relationships between ASD and its related traits [51]. Instrumental strengths, approximated with F1 score > 10 , were calculated using SNP effect and standard error values [6, 49]. To account for the potential biases due to participant overlap between cohorts, the lower bound (95% confidence interval) of the F1 was calculated [9]. The analyses were performed using the platform TwoSampleMR [6].

Whole genome sequencing

The results were validated using yet unpublished data from the EU Horizon2020 GEMMA research project with genotype variant calls in 112 (49% female) WGS samples with 45 ASD probands (42% female) from the GEMMA prospective cohort [70]. These samples, assayed on whole blood and collected during enrollment, were sequenced with 30-40X coverage on Illumina NovaSeq 6000 platform. Data was aligned to GRCh38 reference genome using bwa mem v0.7.17 [34] and reads were sorted and duplicates marked with samtools v1.12 [35]. Quality control was performed with omnomicsQ -software [20]. For variant calling DeepVariant v1.4.0 [54] was utilized and variants were annotated with Variant Effect Predictor [43] version 112.0.

Statistical analysis

All statistical analyses were performed using R 4.2.2 software and available as R markdown results in the github project (https://github.com/jakelin212/mvasd_gwas). Genome-wide association is called on the standard and strict p-value threshold of $5e-8$ ($-\log_{10} 7.3$), to account for multiple testing based on the assumption of about 1-million independent tests [56]. To assess SNP allele proportional differences for validation, the phi coefficient is computed, and statistical significance was determined using Chi-square test. Fisher's exact test was used when Chi-square assumptions were not met. Bonferroni correction is assessed to account for multiple testing of the multivariate GWAS involving 9 traits ($p < 5.5e-9$; $-\log_{10}(p) > 8.25$).

Results

GWAS summary statistics

GWAS summary statistics for ASD and ADHD were collected from the PGC and iPSYCH [49, 65] studies. Education attainment [47] summary file was collected from the Social Science Genetic Association Consortium (SSGAC). Altogether, using summary statistics, 11 ASD co-occurring traits were assessed for genetic correlation with the landmark ASD study [19], the largest genetic correlation values, as computed by LDSC [8], were between ASD and ADHD ($rg = 0.535$), followed by MDD ($rg = 0.505$) and ADHD childhood ($rg = 0.478$). Shown in Table 1 below, 8 traits are shown to be genetically correlated with ASD ($p < 0.05$) and additional details of all traits are shown in Supplementary Table 1.

Multivariate ASD central trait SNPs, pathway and organ tissue enrichment

Multivariate GWAS was performed with ASD together with its genetically correlated traits, ADHD, ADHD childhood, ASRD, bipolar, DBD, EA, MDD, and SCZ (Table 1) and 637 ($p < 5e-08$) SNP associations were found, including 322 variants that are reported for the first time for any trait (Supplementary Table 6) according to GWAS catalog. Two associations (rs2388334 and rs1452075) intersected with the twelve associations identified in the landmark common genetic variants of ASD study [19]. When assessed at the gene level, all 12 were concordant (as indicated in STable 6). Decomposition implemented in MetaPhat, using stepwise tracing of p-value and Bayesian information criteria (BIC) contributions [36, 62], identified 37 ASD central trait SNPs where 16 were identified with multivariate GWAS approach (all SNPs $p < 5.5e-09$; min ($-\log_{10}(p)$) 8.67), listed in Supplementary Table 2). These 37 multivariate ASD SNPs, 17 of which had previously been reported in existing GWAS studies, mapped to 35 genes (Table 2) and confirmed that 8/35 (*ARHGAP32*, *CADPS*, *CUL3*, *KANSL1*, *MACROD2*, *MAPT*, *MSRA* and *NEGR1*) are known curated SFARI genes, with ASD susceptibility EAGLE scores ≤ 3 (indicating limited evidence) [61]. The variant rs538628 within the *NSF* gene, a regulator of AMPA receptor endocytosis and critical for mediating glutamatergic synaptic transmission [25], along with the variant rs62061734 in the *MAPT* gene, are identified to associate with the optimal central traits of ASD, EA and SCZ (*MAPT* variant rs62061734 $p = 3.98e-31$, *NSF* variant rs538628 $p = 1.99e-27$, Supplementary Table 2, trace plots are provided in supplementary data). Notably, *NSF* was previously implicated only in mouse models exhibiting ASD-like behaviors [76]. Shown in the same table, MTAG [71] multivariate GWAS validation was

performed to address iPSYCH cohort sample overlaps between ADHD and ASD [40, 41] subjects where similar results were found (*MAPT* variant rs62061734 $p = 1.99e-20$, *NSF* variant rs538628 $p = 5.37e-18$).

Shown in Supplementary Table 7, Fig. 2e and Supplementary Fig. 3, pathway enrichment using the 35 associated genes was performed with Enrichr [31]. Nervous systems development (GO:0007399) was found to be the most significant ($p = 1.73e-08$) while neural and microtubule structural related pathway hits from Reactome [16] and WikiPathways [46] featured pathways were Inclusion Body Myositis (*MAPT* and *PSENI*, $p = 1.27e-04$) and COPII-mediated Vesicle Transport (*NSF* and *SERPINA1*, $p = 4.69e-03$). Enrichment analysis was conducted using the WebCSEA tool, which identified statistically significant associations (Fig. 2f, $p < 1e-03$) with the following human organ systems: digestive, nervous, sensory, lymphatic, and respiratory. As shown in Supplementary Fig. 4, the most enriched tissue types are related to cerebrum, cortex, intestine and blood related components discerned from 1,355 tissue-type (TS) as well as data from the human brain single cell project [33].

Colocalization analyses

Colocalization analysis was conducted on the 37 multivariate SNP associations identified to contain ASD 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 regulating and promoter regional elements. For the two SNPs that did not map to a gene, the window size used for the colocalization analysis was 100 KB (± 50 KB), estimated and derived from the gene median length of 24KB [17]. Additional information concerning the number of regional LD adjusted SNPs applied to the colocalization test is shown in Supplementary Table 3.

A total of 19/37 SNPs showed strong evidence for a common liability variant with ASD (H4 > 90%, details shown in Supplementary Table 3) and the traits having common ASD liable variants included EA (9), SCZ (6), BP (2), ADHD (1) and ASRD (1). Notably, SNP rs62061734, mapping to the *MAPT* gene and rs538628,

mapping to the *NSF* gene had H4 of 99% for EA and SCZ, respectively (shown in Fig. 2a-b) while SNP rs568828, mapping to the *NTM* gene had H4 of 99% for SCZ and ADHD (Supplementary Table 2).

Mendelian randomization analyses

Mendelian randomization analysis was conducted for the 8 traits genetically correlated (Table 1) with ASD. The lead SNPs, with F1 scores > 25 (listed in Supplementary Table 4, where > 10 is considered strong [48]) were found to lend significantly increase probability of ASD ($p < 0.001$) both Inverse Variance Weighted (IVW)-method and MR-Egger (EA and SCZ are shown in Fig. 2c-d), accounting for horizontal pleiotropy and multiple testing with Bonferroni correction of 8 traits). Based on TwoSampleMR Steiger [22] test for directionality and shown in supplementary Table 4 A, genetic liability to ADHD childhood ($p < 2.44e-116$), ASDR ($p < 9.08e-166$) and DBD ($p < 1.20e-45$) were found to have causal effects on the risk of ASD. Shown in Supplementary Table 4B, genetic liability to ASD ($p < 4.1e-115$) were found to have causal effects on the risk of ADHD, ADHD childhood, BIP, EA, MDD and SCZ. The related MR results adhere to the MR-STROBE guidelines [64].

Validation

To assess the impact of the reported multivariate associations on expression (eQTL) and splicing regulatory quantitative trait loci across tissues, the majority (22/37 eQTL, 24/37 sQTL, details listed in Supplementary Table 9) of the associations found are cited in the EBI QTL Catalog [28] where they associate (adjusted $p < 0.05$) with adipose, brain and neuron tissues. Furthermore, filtering on GeneCards [60] curations, the presented ASD central genes are enriched with systems related to gut, microbiome, intestinal immune, enteric nervous and central nervous systems (Supplementary Table 5).

Additionally, the distribution of these ASD-central trait related SNPs in 112 (49% females; 45 ASD probands (42% females) prospective from-birth GEMMA [70] cohort participants was investigated. SNP distribution differences were for variant rs568828, mapped to *NTM* and

(See figure on next page.)

Fig. 2 Results from the post GWAS analysis of the 37 selected SNPs. **a,b**) Colocalization processing using the original summary statistics of ASD and EA for (a) rs62061734 (*MAPT*, failed colocalization with H4 probability 8.19%, $p = 0.09$), ASD and *NSF* for (b) rs538628 (*NSF*, SCZ passed colocalization with H4 probability 94%, $p = 1.1e-05$), depicting supporting regional SNPs (x-axis) and their negative \log_{10} 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 ASD SNP effects (y-axis) and c) EA and d) SCZ effects (x-axis). **e**) Pathway analysis for the genes associated with the selected SNPs shows 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 37 multivariate gene associations and found enrichment ($p < 1e-03$) with the ASD relevant digestive, nervous and sensory organ systems as well as lymphatic and respiratory systems

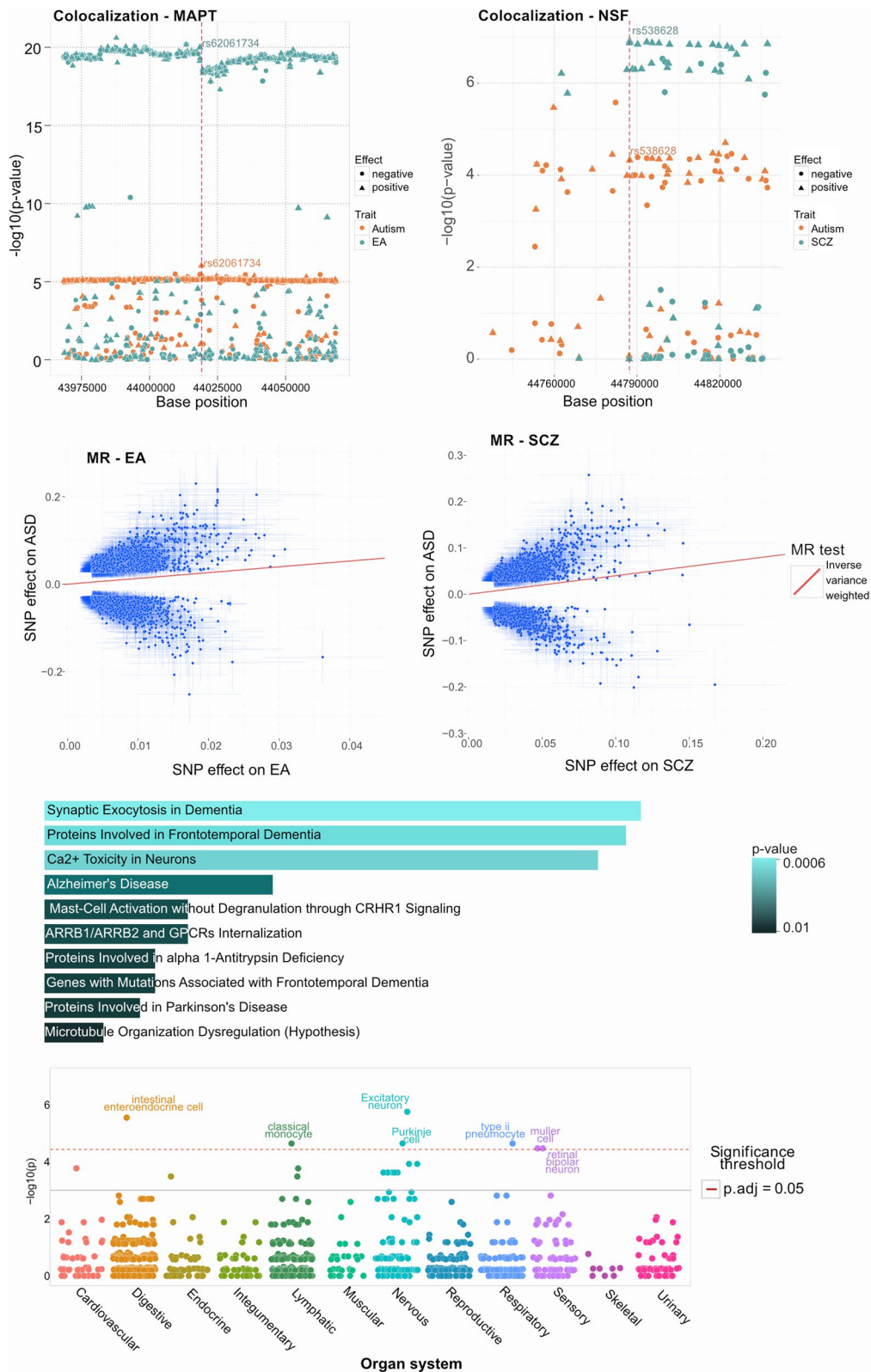


Fig. 2 (See legend on previous page.)

Table 2 Multivariate GWAS ASD-central SNPs tested with coloc and MR tests towards the identified ASD central traits, with all 8 traits passing MR and 19 gene regions/traits pairings passed Coloc ($H4 > 90\%$), indicated with †. Coloc and MR details are additionally listed in supplementary tables 3 and 4. The order of the central traits are determined by p -value importance during decomposition processing. Known GWAS associations (17/37) are marked as * while SFARI ASD gene members (8) are in bold

Rsid	Gene	Chr: pos: ref > alt	SNP consequence	Central traits
rs6699841	NEGR1	1:72645850:A > G	intergenic	EA, ASD
rs67980110	ENSG00000237435	1:96470851:T > C	regulatory	EA, ASD, ADHD
rs2391769*	NA	1:96978961:A > G	intergenic	ASD, ADHD
rs58378462+	ENSG00000221849	2:104138639:A > G	intron	ASD, EA+, ADHDCHILD, ADHD
rs11897599	MRPS18BP2	2:140449566:A > G	downstream	EA, ASD, ADHD
rs78826721+	NDUFS1	2:207002314:A > G	intron	SCZ+, EA, ASD
rs6748341*	CUL3	2:225377574:T > C	regulatory	ASD, EA, SCZ
rs1452075*	CADPS	3:62481063:T > C	intron	EA, ASD
rs62243489+	CADPS	3:62482927:T > G	intron	EA+, ASD
rs6806355+	NA	3:70488292:T > G	intergenic	BP, SCZ+, EA+, ASD
rs35544582+	SLC30A9	4:42044036:A > C	intron	EA+, ASD
rs67779882	NA	5:92488009:A > G	intron, non_coding	EA, ASD, ADHD
rs406413*	ENSG00000246316	5:113898581:T > C	intron	EA, ASD
rs2388334*	ENSG00000271860	6:98591622:A > G	intron	BP, EA, ASD
rs6999466+*	MSRA	8:10265712:A > G	intron	MDD, EA+, ASD
rs877116*	ENSG00000253695	8:10712945:T > G	intron	ASD, ASRD
rs2409743*	ENSG00000270076	8:11070360:T > G	intron, non_coding	ASD, ASRD
rs2409784+	BLK	8:11396856:A > C	intergenic	ASD, ASRD+
rs11775333	NA	8:142637867:T > C	regulatory	EA, ASD
rs11143599	ENSG00000221844	9:76101777:T > G	intron, non_coding	BP, ASD
rs1848797+*	ENSG00000238280	10:64552934:A > G	intron	BP+, EA, ASD
rs12761761*	BNIP3	10:133775375:T > C	downstream	EA, ASD
rs2237943	SERGEF	11:17838248:T > C	regulatory	EA, ASD, ADHD
rs4609618	ARHGAP32	11:128818792:A > C	intergenic	ASD, ADHD
rs568828+	NTM	11:131732259:T > G	intergenic	ASD, SCZ+, ADHD+
rs177413+	PSEN1	14:73683194:T > C	intergenic	BP+, ASD, EA, SCZ
rs736281*	NA	14:94287830:T > C	intron	EA, ASD
rs28929474*	SERPINA1	14:94844947:T > C	intron	EA, ASD
rs62065453+*	ENSG00000131484	17:43573419:A > G	regulatory	ASD, EA+, SCZ+
rs62057107+*	CRHR1	17:43896032:T > C	intergenic	ASD, EA+, SCZ+
rs62061734+*	MAPT	17:44018488:T > C	intron	ASD, EA+, SCZ
rs2696633+	KANSL1	17:44270059:T > G	intron	ASD, EA+, SCZ
rs538628*	NSF	17:44787313:T > C	regulatory	ASD, EA, SCZ+
rs1792709*	ENSG00000206129	18:53768975:A > G	intron	ASD, SCZ
rs6079546	MACROD2	20:14716738:T > G	intergenic	MDD, EA, ASD
rs6035835	XRN2	20:21271669:A > G	intergenic	ASD, ADHDCHILD, ADHD
rs9974470	ENSG00000249209	21:35012066:A > G	intron	EA, ASD

rs62243489, mapped to *CADPS*. The *NTM* gene, previously associated with emotional learning deficits in murine models [42], encodes neurotrimin, while *CADPS* encodes a neural/endocrine-specific membrane protein regulating calcium. The *NTM* SNP (rs568828) was present in 42 of 45 probands (92%) compared to 100% of controls (67 of 67). In contrast, *CADPS* SNP (rs62243489) was found in 19/67 controls (28%) and enriched in 21/45

probands (47%). As listed in Supplementary Table 8, the phi coefficient for *NTM* between probands and controls was 0.2 ($p=0.062$), while for *CADPS*, it was -0.19 ($p=0.047$). When stratified by sex, the phi coefficient for *NTM* in males was 0.15 ($p=0.456$) and 0.27 ($p=0.040$) in females. For *CADPS*, the phi coefficients were -0.18 ($p=0.182$) in males and -0.20 ($p=0.140$) in females. Notably, the *NEGR1* gene (variant rs6699841), involved

in neuron growth regulation, showed a phi coefficient of -0.27 ($p=0.040$) in males (24/26 cases; 22/31 controls), while in females, the coefficient was 0.26 ($p=0.084$; 12/19 cases; 31/36 controls). For the variant of *NEGR1*, the opposing phi directions between sexes resulted in a phi coefficient of -0.01 ($p=0.908$) in the full dataset. In addition, logistic regression was performed for the specific variants of *NTM*, *CADPS* and *NEGR1*. The results were not significant for the full cohort (adjusted for sex) or in models stratified by sex.

Discussion

Using multivariate statistical learning approaches, this study constitutes the largest and most comprehensive genetically correlated multi-trait GWAS analysis with summary statistics performed on ASD and its genetically correlated traits; ADHD, ADHD childhood, ASRD, bipolar, DBD, EA, MDD, and SCZ to explore the underpinnings driving the complexities in ASD. 37 associations containing ASD as a central trait were discovered, with 16 of these associations were detected only due to the increased statistical power of this multivariate GWAS analysis (lowest univariate summary statistics p-value from all traits $>5e-08$, and 12/16 confirmed with the MTAG tool [71], Supplementary Table 2). Interestingly, a previous study using electronic health records of covering nearly 5,000 ASD cases found three subclusters of comorbidity trajectories, first characterized by seizures, then auditory disorders/infections and the third cluster by psychiatric disorders. Due to the complexity of ASD development, a fourth group was described as could not be further resolved. The presented subclusters potentially align well with our ASD central trait sets pertaining to SCZ signals with seizures and psychiatric disorders such as ADHD and intellectual development underpinning EA [15]. Enrichment analysis confirmed that the multivariate ASD association results are related to neuron and gut tissues and developmental pathways as well as inflammation and microbiome domains, further underscoring the intersection of genome and microbiome as well as supportive of the gut-brain axis hypothesis associated to ASD [11, 44]. Surprisingly, genetic correlation performed on LDSC indicated that ASD and IBD are not related (Supplementary Table S1, $rg = -0.059$; $p=0.44$), a recent report highlighted potential evidence for comorbidity between parental, particularly maternal preexisting IBD onsets and their children developing ASD [59]. Using the multivariate ASD central trait gene sets, based on comprehensive human tissue cell type and single cell data [13, 33] analysis, enrichments were detected with digestive, nervous, and sensory organ systems (Fig. 2f). At the tissue cell type level and

further supporting the gut-brain axis and blood brain barrier, the analysis detected enriched ASD relevant signals related to brain, adipose and gut eQTL/sQTL (Supplementary Table 9) tissue panels.

Overall, the identified ASD traits passed MR with strong F1 measures and significantly contributed to improve the future construction of meta psychiatric based ASD polygenic scores [27], shown to improve prediction relative to standard PRS in other complex traits such as coronary heart disease and type 2 diabetes [37, 67]. The MR results were consistent after calculation of lower bound F1 (all scores >25 , Supplementary Table 4), computed to consider potential biases from cohort sample overlaps [9]. MR Steiger tests for directionality revealed that genetic liability to ASDR and DBD were found to have causal effects on the risk of ASD. These multivariate ASD associations mapped to genes, including *MAPT* and *NSF* which are known to be involved in biological pathways linked to neural disorders such as infantile epilepsy [66] and Parkinson's Disease [7, 14]. Interestingly, colocalization tests for the *MAPT* region indicated shared genetic risk between only EA and ASD (H4 99%), while that for the *NSF* gene did not associate with EA, instead associated with SCZ (H4 94%), suggesting intra region heterogeneity that demands future investigation. With respect to ASD, the *KANSL1*, *BNIP3*, *CADPS* and *NEGR1* genes have been implicated with immune and microbiome features [11] and behavioral developments [63]. Similarly, a recent study from Arenella and colleagues reported genetic factors between ASD and various immune phenotypes including *KANSL1* associating with lymphocyte counts as well as *MAPT* associating with eosinophil counts, further supporting the role of the inflammation pathway in ASD development [1].

The most common traits in our set of 37 associations that passed colocalization with ASD were EA (9), SCZ (6) and BP (2). It is known that the diagnosis for ASD and ADHD, particularly ADHD manifestation in young children, is similar with symptomatic issues concerning hyperactivity and attention span [29]. While a previous study has performed comparison of genetic and functional enrichment of associations between ASD and ADHD [50] GWAS resources, this study further complements their results by inclusion of other ASD co-occurring traits, including ADHD and ADHD childhood as well as EA. Interestingly, ASD and ADHD have both been linked with dysbiosis disruption in microbiome composition and function, gastrointestinal and bowel habits issues [44].

As part of validation, clustering and distribution proportion differences based on the ASD identified SNP associations were detected between probands and

non-autistic subjects on genomes from the GEMMA cohort [70]. Our validation results were performed on the (112, 45 ASD probands) samples currently available in GEMMA. Notably, *NEGR1* (rs1432639), a neuronal growth regulator known to associate with migraine, depression and seizures [24, 26, 73], the significant phi coefficients were negative for males and positive for females. Interestingly, a previous study pertaining to prenatal stress found increased *NEGR1* expression in the hippocampus of female rats but not in males [72]. To improve on the specificity and clinical value of the identified traits, a follow-up application of MR with specific expression/protein quantitative loci (tissue/cell type e/pQTL as applied in T1D drug candidate discovery [21] with genes such as *CADPS*, *NTM* and *NEGR1* could further reveal molecular and translational insights towards ASD heterogeneity including the high vulnerability subgroup characterized by seizures [15]. While the validation statistical power was limited by the relatively small sample size, nevertheless the independent and deep sequencing data has allowed the harvesting of interesting observations concerning the distribution of ASD-central trait associations in probands as compared to controls. In addition, the GEMMA validation results should be taken with caution as the population structures (PCs) were not included as covariates due to availability. The upcoming release of additional omics data from GEMMA and other studies, including longitudinal microbiome, metabolome, and methylation datasets, will significantly increase statistical power and enable more detailed temporal analyses. The data will help confirm molecular changes along the gut-brain axis, shedding light on the genetic patterns that contribute to the heterogeneity, development, and comorbidities of ASD. Another limitation of our multi-trait GWAS is that the selection of ASD co-occurring traits is not exhaustive; given the complexity of ASD development, there may be other genetically correlated traits that have not yet been tested at the appropriate population level, warranting consideration and inclusion in future studies. The MR Steiger results on causality need to be taken with caution as unmeasured confounding effects may distort the exposure genetic liability relative to the outcome [38].

Conclusion

Our study represents the largest multivariate GWAS on ASD to date, combining ASD with eight genetically correlated trait GWAS summaries. We performed systematic decomposition to identify novel genetic associations related to ASD and ASD co-occurring traits. Mendelian randomization testing revealed that genetic liability for ADHD childhood, ASRD and DBD has causal effects

on the risk of ASD. Colocalization analysis further confirmed shared genetic risks with ASD, showing enrichment patterns in brain tissues and cell types associated with neurodevelopment, and lending additional support to the gut-brain axis hypothesis.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12888-024-06392-w>.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Figure 1. Genetic correlation of the traits included in the analysis. ASD = Autism Spectrum Disorder; ADHD = Attention Deficit Hyper Disorder; ASRD = Anxiety-Stress Disorder; DBD = Disruptive Behaviour Disorder; EA = Education attainment; MDD = Major Depression Disorder; SCZ = Schizophrenia. Supplementary Figure 2. *P*-value and BIC decomposition processing of MAPT and NSF to identify ASD central traits. ASD = Autism Spectrum Disorder; ADHD = Attention Deficit Hyper Disorder; ASRD = Anxiety-Stress Disorder; DBD = Disruptive Behaviour Disorder; EA = Education attainment; MDD = Major Depression Disorder; SCZ = Schizophrenia. Supplementary Figure 3. Pathway analysis using the WikiPathway database also highlights neuronal processes, with bar length and color indicating significance. More details listed in Supplementary Table 8. Supplementary Figure 4. Tissue and cell (TS) type enrichment using WebCSEA and the list of the 22 central trait genes found that the most enriched tissues are related to cerebrum, cortex and small intestine related tissue types. Lake 2017 refers to data from human brain single cell analysis project (<https://pubmed.ncbi.nlm.nih.gov/29227469/>) while HCA stands for histologic chorioamnionitis, an intrauterine inflammatory trait. Supplementary Figure 5. ASD multivariate GWAS associations within the MAPT H1/H2 haplotype, 17q21 arm region, are presented in a Manhattan plot, in the context of Grove et al. GWAS results. Significance thresholds for *p*-values of $1e-05$ indicated in blue and $1e-08$ in red. Significant SNPs highlighted in green show rs62061734 (MAPT), rs269633 (KANS1) and rs538628 (NSF).

Supplementary Table 1. Data and sample details of ASD and 8 genetically correlated traits ($P < 0.05$, calculated from LD Score Regression (LDSC)) are presented and applied towards multivariate-GWAS. Data from four excluded traits are additionally shown. Supplementary Table 2. 37 multivariate associations are identified with ASD as a central trait where 17/37, shown with asterisk are previously reported in the GWAS Catalog and in bold, 8 genes are identified as SFARI ASD genes. Supplementary Table 3. 19 gene regions/trait pairings passed coloc (Posterior Prob $H4 > 90\%$, Shown in bold) called on coloc.abf with a window size of ± 50 KB flanking the SNP locus. Supplementary Table 4. (A) Mendelian randomization (MR) results for ASD as outcome and related traits. (B) MR where ASD is the exposure and related traits are the outcome. Supplementary Table 5. MV associated genes are found in systems curated/implicated with gut microbiome and neural systems from GeneCards. Supplementary Table 6. List of 637 Significant SNPs ($p < 5e-8$), with 315 already reported in the GWAS catalog, identified by MetaPhat multivariate-GWAS using ASD and 8 genetically correlated trait summary statistics. Supplementary Table 7. A) 108 enriched ($p < 0.05$) Go terms are annotated and (B) 46 pathways on WikiPathway C) KEGG D) Reactome resources e) Tissue from the list of multivariate ASD SNPs found enrichments in neuron and nervous systems related data. Supplementary Table 8. ASD central SNP alleles are mapped to GEMMA genotypes called from 112 (49% females) WGS samples (45 (42% females) ASD probands). Phi coefficients are calculated between allele proportions where Chi-square test is applied to assess statistical importance. Indicated with *. Fisher's exact test is applied when Chi-square assumptions are violated. Supplementary Table 9. eQTL and sQTL related results of the ASD central associations relative to brain and nervous systems from EBI QTL catalog are captured via <https://fivex.sph.umich.edu/>. Study URLs are listed at the bottom of the table.

Acknowledgements

The work was supported by the European Commission Horizon 2020 programme (grant no. 825033) and the Strategic profiling of Tampere University in health data science, Academy of Finland, PROFI 6, 2021–2026. We are sincerely grateful to the entire GEMMA team and in particular to the nurses, recruiters and especially families participating in the project.

Authors' contributions

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

Funding

Open access funding provided by Tampere University (including Tampere University Hospital). The work was supported by the GEMMA project, funded by the European Commission by means of the Horizon 2020 program (call H2020-SC1-BHC-03-2018) with the project ID 825033 and the Strategic profiling of Tampere University in health data science, Academy of Finland, PROFI 6, 2021–2026.

Data availability

All data generated or analyzed during the study are included and additionally available upon request. For scripts, please see: https://github.com/jakelin212/mvasd_gwas

Declarations

Ethics approval and consent to participate

The GEMMA study was approved by the relevant ethics committee of each enrolling country. Particularly, CE Campania Sud (IRB n.30/2019) for Italy; Partners Human Research (IRB ver.01/04/2019) for USA; and Clinical Research Ethics Committee of Galway University Hospital (IRB n. C.A. 2127/19) for Ireland. A written consent form will be signed by each participant or their legal representative.

Consent for publication

Consent, relevant to GEMMA subjects, is granted and signed by each participant or their legal representative.

Competing interests

The authors declare no competing interests.

Author details

¹Faculty of Medicine and Health Technology, Tampere University and Tays Cancer Centre, Tampere, Finland. ²The GEMMA study, GEMMA-PROJECT.eu, Salerno, Italy. ³BMSsystems, Paris, France. ⁴Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, USA. ⁵Eufomatics, Tekniikantie, Espoo, Finland. ⁶Foundation for the Finnish Cancer Institute, Helsinki, Finland. ⁷European Biomedical Research Institute of Salerno (EBRIS), Salerno, Italy. ⁸Harvard Medical School, Harvard T.H. Chan School of Public Health, Boston, USA. ⁹Health Sciences, Faculty of Social Sciences, Tampere University, Tampere, Finland. ¹⁰Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden.

Received: 1 August 2024 Accepted: 8 December 2024

Published online: 18 December 2024

References

- Arenella M, Fanelli G, Kiemeny LA, Grainne McAlonan DG, Murphy, and Janita Bralten., *Brain., behavior, & immunity - health* vol. 34 100698. 3 Nov, <https://doi.org/10.1016/j.bbih.2023.100698>
- Arpi MN, Torres, Ian Simpson T. SFARI genes and where to find them; modelling Autism Spectrum Disorder Specific Gene expression dysregulation with RNA-Seq Data. *Sci Rep.* 2022;12(1):10158.
- Ashburner J, Ziviani J, and Sylvia Rodger. Surviving in the mainstream: capacity of children with Autism Spectrum disorders to perform academically and regulate their emotions and behavior at School. *Res Autism Spectr Disorders.* 2010;4(1):18–27.
- Baio J, Wiggins L, Christensen DL, Maenner MJ, Daniels J, Warren Z, Kurzius-Spencer M, et al. Prevalence of Autism Spectrum Disorder among children aged 8 years - Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2014. *Morbidity Mortal Wkly Rep Surveillance Summaries.* 2018;67(6):1–23.
- Boorstein HC. Gastrointestinal and feeding problems in Young children: a comparison between Autism Spectrum disorders and other Developmental delays. University of Connecticut; 2008.
- Bowden J, Smith GD, and Stephen Burgess. Mendelian randomization with Invalid instruments: Effect Estimation and Bias Detection through Egger Regression. *Int J Epidemiol.* 2015;44(2):512–25.
- Brion JP, Octave JN, Couck AM. Distribution of the Phosphorylated Microtubule-Associated protein tau in developing cortical neurons. *Neuroscience.* 1994;63(3):895–909.
- Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson N, Daly MJ, Price AL, Neale BM. LD score regression distinguishes confounding from polygenicity in genome-wide Association studies. *Nat Genet.* 2015;47(3):291–95.
- Burgess S, Davies NM, and Simon G. Thompson. Bias due to participant overlap in two-sample mendelian randomization: Burgess Et Al. *Genet Epidemiol.* 2016;40(7):597–608.
- Chaste P, and Marion Leboyer. Autism risk factors: genes, Environment, and gene-environment interactions. *Dialog Clin Neurosci.* 2012;14(3):281–92.
- Cheng L, Wu H, Chen Z, Hao H, and Xiao Zheng. Gut microbiome at the crossroad of genetic variants and Behavior disorders. *Gut Microbes.* 2023;15(1):2201156.
- Cichonska A, Rousu J, Marttinen P, Kangas AJ, Soininen P, Lehtimäki T, Raitakari OT, et al. metaCCA: Summary statistics-based multivariate Meta-analysis of Genome-Wide Association Studies Using Canonical Correlation Analysis. *Bioinformatics.* 2016;32(13):1981–89.
- Dai Y, Hu R, Liu A, et al. WebCSEA: web-based cell-type-specific enrichment analysis of genes. *Nucleic Acids Res.* 2022;50(W1):W782–90. <https://doi.org/10.1093/nar/gkac392>.
- Derkinderen P, Rolli-Derkinderen M, Chapelet G, Neunlist M, and Wendy Noble. Tau in the gut, does it really Matter? *J Neurochem.* 2021;158(2):94–104.
- Doshi-Velez, Finale Y, Ge, and Isaac Kohane. Comorbidity clusters in Autism Spectrum disorders: an Electronic Health Record Time-Series Analysis. *Pediatrics.* 2014;133(1):e54–63.
- Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, Haw R, et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* 2018;46(D1):D649–55.
- Fuchs G, Voicheck Y, Benjamin S, Gilad S, Amit I, and Moshe Oren. 4sUDRB-Seq: measuring Genomewide Transcriptional Elongation Rates and initiation frequencies within cells. *Genome Biol.* 2014;15(5):R69.
- Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, Milind Mahajan, et al. Most genetic risk for Autism resides with common variation. *Nat Genet.* 2014;46(8):881–85.
- Grove J, Ripke S, Als TD, Manuel Mattheisen RK, Walters H, Won J Pallesen, et al. Identification of Common Genetic Risk variants for Autism Spectrum Disorder. *Nat Genet.* 2019;51(3):431–44.
- Gutowska-Ding M, Weronika ZC, Deans C, Roos J, Matilainen F, Khawaja K, Brügger JW, Ahn C, Boustred, and Simon J. Patton. One byte at a time: evidencing the quality of clinical service next-generation sequencing for germline and somatic variants. *Eur J Hum Genetics: EJHG.* 2020;28(2):202–12.
- Heikkilä TE, Emilia K, Kaiser J, Lin D, Gill JJ, Koskeniemi. and Ville Karhunen. 2024. Genetic Evidence for Efficacy of Targeting IL-2, IL-6 and TYK2 Signalling in the Prevention of Type 1 Diabetes: A Mendelian Randomisation Study. *Diabetologia*, September. <https://doi.org/10.1007/s00125-024-06267-5>
- Hemani G, Tilling K, George Davey Smith. Orienting the causal relationship between Imprecisely measured traits using GWAS Summary Data. *PLoS Genet.* 2017;13(11):e1007081.
- Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, Tan VY, Yarmolinsky J, Shihab HA, Timpson NJ, Evans DM, Relton C, Martin RM, Davey Smith G, Gaunt TR,

- Haycock PC. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7:e34408. <https://doi.org/10.7554/eLife.34408>. PMID: 29846171; PMCID: PMC5976434.
24. Hyde CL, Nagle MW, Tian C, Chen X, Paciga SA, Wendland JR, Joyce Y, Tung DA, Hinds RH, Perlis, and Ashley R. Winslow. Identification of 15 genetic loci Associated with risk of Major Depression in individuals of European descent. *Nat Genet*. 2016;48(9):1031–36.
 25. Iwata K, Matsuzaki H, Tachibana T, Ohno K, Yoshimura S, Takamura H, Yamada K, et al. N-Ethylmaleimide-sensitive factor interacts with the Serotonin Transporter and modulates its trafficking: implications for pathophysiology in Autism. *Mol Autism*. 2014;5:33.
 26. Jankowsky JL, Patterson PH. The role of cytokines and growth factors in seizures and their sequelae. *Prog Neurobiol*. 2001;63(2):125–49.
 27. Jansen AG, Gwen C, Dieleman PR, Jansen FC, Verhulst D, Posthuma, Tinca JC, Polderman. Psychiatric Polygenic Risk scores as Predictor for attention Deficit/Hyperactivity Disorder and Autism Spectrum Disorder in a clinical child and adolescent sample. *Behav Genet*. 2020;50(4):203–12.
 28. Kerimov N, Hayhurst JD, Peikova K, Manning JR, Walter P, Kolberg L, Samoviča M et al. 2021. eQTL Catalogue: A Compendium of Uniformly Processed Human Gene Expression and Splicing QTLs. *bioRxiv*. <https://doi.org/10.1101/2020.01.29.924266>
 29. Kern JK, Geier DA, Sykes LK, Geier MR, Deth RC. Are ASD and ADHD a Continuum? A comparison of pathophysiological similarities between the disorders. *J Atten Disord*. 2015;19(9):805–27.
 30. Khachadourian V, Mahjani B, Sandin S, Kolevzon A, Buxbaum JD. Abraham Reichenberg, and Magdalena Janecka. 2023. Comorbidities in Autism Spectrum Disorder and their etiologies. *Translational Psychiatry* 13 (1): 71.
 31. Kuleshov MV, Jones MR, Rouillard AD, Nicolaf F, Fernandez Q, Duan Z, Wang S Koplev, et al. Enrichr: a Comprehensive Gene Set Enrichment Analysis web server 2016 Update. *Nucleic Acids Res*. 2016;44(W1):W90–97.
 32. Kwong A, Boughton AP, Wang M, VandeHaar P, Boehnke M. Gonçalo Abecasis, and Hyun Min Kang. 2022. FIVEx: an interactive eQTL browser across Public Datasets. *Bioinformatics* 38 (2): 559–61.
 33. Lake BB, Chen S, Sos BC, et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat Biotechnol*. 2018;36(1):70–80. <https://doi.org/10.1038/nbt.4038>.
 34. Li H. 2013. Aligning sequence reads, clone sequences and Assembly contigs with BWA-MEM. <http://arxiv.org/abs/1303.3997>
 35. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, R. and 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics*. 2009;25(16):2078.
 36. Lin J, Tabassum R, Ripatti S, Pirinen M. MetaPhat: detecting and decomposing Multivariate associations from Univariate Genome-Wide Association Statistics. *Front Genet*. 2020;11:431.
 37. Lin J, Mars N, Fu Y, Ripatti P, Kiiskinen T, Tukiaainen T, Ripatti S. and Matti Pirinen. Integration of Biomarker Polygenic Risk Score Improves Prediction of Coronary Heart Disease. *JACC: Basic to Translational Science*. 2023. <https://doi.org/10.1016/j.jacbs.2023.07.006>
 38. Lutz SM, Voorhies K, Wu AC, Hokanson J, Vansteelandt S, and Christoph Lange. The influence of unmeasured confounding on the MR Steiger Approach. *Genet Epidemiol*. 2022;46(2):139–41.
 39. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, et al. The New NHGRI-EBI Catalog of Published Genome-Wide Association Studies (GWAS catalog). *Nucleic Acids Res*. 2017;45(D1):D896–901.
 40. Mattheisen M, Grove J, Als TD, Martin J, Voloudakis G, Meier S, Demontis D, et al. Identification of Shared and differentiating Genetic Architecture for Autism Spectrum Disorder, attention-deficit hyperactivity disorder and case subgroups. *Nat Genet*. 2022a;54(10):1470–78.
 41. Mattheisen M, Grove J, Als TD, Martin J, Voloudakis G, Meier S, Demontis D, Bendl J, Walters R, Carey CE, Rosengren A, Strom NI, Hauberg ME, Zeng B, Hoffman G, Zhang W, Bybjerg-Grauholm J, Bækvad-Hansen M, Agerbo E, Corman B, Nordentoft M, Werge T, Mors O, Hougaard DM, Buxbaum JD, Faraone SV, Franke B, Dalsgaard S, Mortensen PB, Robinson EB, Roussos P, Neale BM, Daly MJ, Børglum AD. Identification of shared and differentiating genetic architecture for autism spectrum disorder, attention-deficit hyperactivity disorder and case subgroups. *Nat Genet*. 2022b;54(10):1470–8. <https://doi.org/10.1038/s41588-022-01171-3>. Epub 2022 Sep 26. PMID: 36163277; PMCID: PMC10848300.
 42. Mazitov T, Bregin A, Philips M-A, Innos Jürgen, and Eero Vasar. Deficit in emotional learning in Neurotrimin Knockout Mice. *Behav Brain Res*. 2017;317(January):311–18.
 43. McLaren W, Gil L, Hunt SE, Riat HS, Graham RS, Ritchie. Anja Thormann, Paul Flicek, and Fiona Cunningham. 2016. The Ensembl variant effect predictor. *Genome Biol* 17 (1): 1–14.
 44. Morton JT, Dong-Min Jin RH, Mills Y, Shao G, Rahman D, McDonald Q, Zhu, et al. Multi-level analysis of the gut-brain Axis shows Autism Spectrum Disorder-Associated Molecular and Microbial profiles. *Nat Neurosci*. 2023;26(7):1208–17.
 45. Nayar K, Sealock JM, Maltman N, Bush L, Cook EH, Lea K, Davis, and Molly Losh. Elevated polygenic burden for Autism Spectrum disorder is Associated with the broad autism phenotype in mothers of individuals with Autism Spectrum Disorder. *Biol Psychiatry*. 2021;89(5):476–85.
 46. Nesterova AP, Yuryev A, Klimov EA, Zharkova M, Shkrob M, Ivanikova NV, Sergey Sozin, and Vladimir Sobolev. 2019. *Disease pathways: an Atlas of Human Disease Signaling pathways*. Elsevier.
 47. Okbay A, Wu Y, Wang N, Jayashankar H, Bennett M, Nehzati SM, Sidorenko J, et al. Polygenic prediction of Educational Attainment within and between families from Genome-Wide Association Analyses in 3 million individuals. *Nat Genet*. 2022;54(4):437–49.
 48. Palmer TM, Debbie A, Lawlor RM, Harbord NA, Sheehan JH, Tobias, Nicholas J, Timpon GD, Smith, Jonathan AC, Sterne. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat Methods Med Res*. 2012;21(3):223–42.
 49. Pedersen CB, Bybjerg-Grauholm J, Pedersen MG, Grove J, Agerbo E, Bækvad-Hansen M, Poulsen JB, et al. The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and Environmental Architectures of severe Mental disorders. *Mol Psychiatry*. 2018;23(1):6–14.
 50. Peyre H, Schoeler T, Liu C, Michèle C, Williams N, Hoertel A, Havdahl, Jean-Baptiste P. Combining Multivariate genomic approaches to elucidate the comorbidity between Autism Spectrum disorder and attention deficit hyperactivity disorder. *J Child Psychol Psychiatry Allied Discip*. 2021;62(11):1285–96.
 51. Phillips AN, Smith GD. How independent are 'Independent' effects? Relative risk estimation when correlated exposures are measured imprecisely. *J Clin Epidemiol*. 1991;44(11):1223–31.
 52. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, Thiruvahindrapuram B, et al. Convergence of genes and Cellular pathways Dysregulated in Autism Spectrum disorders. *Am J Hum Genet*. 2014;94(5):677–94.
 53. Piovesan A, Caracausi M, Antonaros F, Pelleri MC, Vitale L. GeneBase 1.1: A Tool to Summarize Data from NCBI Gene Datasets and Its Application to an Update of Human Gene Statistics. *Database: The Journal of Biological Databases and Curation* 2016. 2016. <https://doi.org/10.1093/database/baw153>
 54. Poplin R, Chang P-C, Alexander D, Schwartz S, Colthurst T, Ku A, Newburger D, et al. A Universal SNP and small-indel variant caller using deep neural networks. *Nat Biotechnol*. 2018;36(10):983–87.
 55. Rees JMB, Angela M, Wood, and Stephen Burgess. Extending the MR-Egger Method for Multivariable mendelian randomization to correct for both measured and unmeasured pleiotropy. *Stat Med*. 2017;36(29):4705–18.
 56. Risch N, Merikangas K. The future of genetic studies of Complex Human diseases. *Science*. 1996;273(5281):1516–17.
 57. Romero M, Aguilar JM, Del-Rey-Mejías Ángel, Mayoral Fermín, Rapado Marta, Peciña Marta, Barbancho Miguel Ángel, Ruiz-Veguilla Miguel, Lara José Pablo. Psychiatric comorbidities in autism spectrum disorder: a comparative study between DSM-IV-TR and DSM-5 diagnosis. *Int J Clin Health Psychol*. 2016;16(3):266–75.
 58. Ruotsalainen SE, Juulia J, Partanen A, Cichonska J, Lin C, Benner I, Surakka FG, et al. An expanded analysis Framework for Multivariate GWAS connects inflammatory biomarkers to functional variants and Disease. *Eur J Hum Genetics: EJHG*. 2021;29(2):309–24.
 59. Sadik A, Dardani C, Pagoni P, Havdahl A, Stergiakouli E, iPSYCH Autism Spectrum Disorder Working Group, Khandaker GM, et al. Parental inflammatory bowel Disease and Autism in Children. *Nat Med*. 2022;28(7):1406–11.
 60. Safran M, Dalah I, Alexander J, Rosen N, Stein TI, Shmoish M, et al. GeneCards Version 3: the human gene integrator. *Database*. 2010;2010:baq020.

61. Schaaf CP, Betancur C, Yuen RKC, Jeremy R, Parr DH, Skuse L, Gallagher RA, Bernier, et al. A Framework for an evidence-based gene list relevant to Autism Spectrum Disorder. *Nat Rev Genet.* 2020;21(6):367–76.
62. Schwarz G. Estimating the dimension of a model. *Ann Stat.* 1978;6(2):461–64.
63. Singh K, Jayaram M, Kaare M, Leidmaa E, Jagomäe T, Heinla I, Hickey MA, et al. Neural cell adhesion Molecule Negr1 Deficiency in Mouse results in structural brain endophenotypes and behavioral deviations related to Psychiatric disorders. *Sci Rep.* 2019;9(1):5457.
64. Skrivankova VW, Richmond RC, Woolf BAR, Davies NM, Swanson SA, VanderWeele TJ, et al. Strengthening the reporting of Observational studies in Epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration. *BMJ.* 2021;375:n2233.
65. Sullivan PF, Agrawal A, Bulik CM, Andreassen OA, Børglum AD, Breen Sven Cichon, et al. Psychiatric genomics: an update and an agenda. *Am J Psychiatry.* 2018;175(1):15–27.
66. Suzuki H, Yoshida T, Morisada N, Uehara T, Kosaki K, Sato K, Matsubara K, Takano-Shimizu T, and Toshiki Takenouchi. De Novo NSF mutations cause early infantile epileptic Encephalopathy. *Ann Clin Transl Neurol.* 2019;6(11):2334–39.
67. Tamlander M, Mars N, Pirinen M, FinnGen E, Widén, and Samuli Ripatti. Integration of Questionnaire-based risk factors improves polygenic risk scores for human coronary heart disease and type 2 diabetes. *Commun Biology.* 2022;5(1):158.
68. Tamm L, Duncan A, Vaughn A, McDade R, Estell N, Birnschein A, and Lori Crosby. Academic needs in Middle School: perspectives of parents and youth with autism. *J Autism Dev Disord.* 2020;50(9):3126–39.
69. Tesi N, van der Lee S, Hulsman M, Holstege H, Marcel JT, Reinders. snpXplorer: a web application to explore human SNP-Associations and annotate SNP-Sets. *Nucleic Acids Res.* 2021;49(W1):W603–12.
70. Troisi J, Autio R, Beopoulos T, Bravaccio C, Carraturo F, Corrivetti G, Cunningham S, et al. 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 Sci.* 2020;10(10). <https://doi.org/10.3390/brainsci10100743>.
71. Turley P, Walters RK, Maghziyan O, Okbay A, Mark Lee JJ, Fontana Alan, Nguyen-Viet Tuan Anh, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet.* 2018;50(2):229–37.
72. Van den Hove DLA, Kenis G, Brass A, Opstelten R, Rutten BPF, Bruschetini M, Blanco CE, Lesch KP, Steinbusch HWM, Prickaerts J. Vulnerability versus resilience to prenatal stress in male and female rats; implications from Gene expression profiles in the Hippocampus and Frontal Cortex. *Eur Neuropsychopharmacology: J Eur Coll Neuropsychopharmacol.* 2013;23(10):1226–46.
73. Vgontzas A, and William Renthal. Migraine-Associated Gene expression in cell types of the Central and Peripheral Nervous System. *Cephalgia: Int J Headache.* 2020;40(5):517–23.
74. Wallace C. Eliciting priors and relaxing the single causal variant Assumption in Colocalisation analyses. *PLoS Genet.* 2020;16(4):e1008720.
75. Wallace C. A more Accurate Method for Colocalisation Analysis allowing for multiple causal variants. *PLoS Genet.* 2021;17(9):e1009440.
76. Xie M-J, Iwata K, Ishikawa Y, Nomura Y, Tani T, Murata K, Fukazawa Y, and Hideo Matsuzaki. Autistic-like Behavior and Impairment of Serotonin Transporter and AMPA receptor trafficking in N-Ethylmaleimide sensitive factor gene-deficient mice. *Front Genet.* 2021;12:748627.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.