

RESEARCH

Open Access



# A family-based study of genetic and epigenetic effects across multiple neurocognitive, motor, social-cognitive and social-behavioral functions

Ron Nudel<sup>1,2</sup>, Richard Zetterberg<sup>3</sup>, Noline Hemager<sup>1,2,4</sup>, Camilla A. J. Christiani<sup>1,2</sup>, Jessica Ohland<sup>1,2</sup>, Birgitte K. Burton<sup>2,4,5</sup>, Aja N. Greve<sup>2,6</sup>, Katrine S. Spang<sup>2,4</sup>, Ditte Ellersgaard<sup>1,2</sup>, Ditte L. Gantriis<sup>2,6</sup>, Jonas Bybjerg-Grauholm<sup>2,7</sup>, Kerstin J. Plessen<sup>2,4,8</sup>, Jens Richardt M. Jepsen<sup>1,2,4,9</sup>, Anne A. E. Thorup<sup>2,4</sup>, Thomas Werge<sup>2,3,5</sup>, Ole Mors<sup>2,6</sup> and Merete Nordentoft<sup>1,2,5\*</sup>

## Abstract

Many psychiatric and neurodevelopmental disorders are known to be heritable, but studies trying to elucidate the genetic architecture of such traits often lag behind studies of somatic traits and diseases. The reasons as to why relatively few genome-wide significant associations have been reported for such traits have to do with the sample sizes needed for the detection of small effects, the difficulty in defining and characterizing the phenotypes, partially due to overlaps in affected underlying domains (which is especially true for cognitive phenotypes), and the complex genetic architectures of the phenotypes, which are not wholly captured in traditional case–control GWAS designs. We aimed to tackle the last two issues by performing GWASs of eight quantitative neurocognitive, motor, social-cognitive and social-behavioral traits, which may be considered endophenotypes for a variety of psychiatric and neurodevelopmental conditions, and for which we employed models capturing both general genetic association and parent-of-origin effects, in a family-based sample comprising 402 children and their parents (mostly family trios). We identified 48 genome-wide significant associations across several traits, of which 3 also survived our strict study-wide quality criteria. We additionally performed a functional annotation of implicated genes, as most of the 48 associations were with variants within protein-coding genes. In total, our study highlighted associations with five genes (*TGM3*, *CACNB4*, *ANKS1B*, *CSMD1* and *SYNE1*) associated with measures of working memory, processing speed and social behavior. Our results thus identify novel associations, including previously unreported parent-of-origin associations with relevant genes, and our top results illustrate new potential gene → endophenotype → disorder pathways.

**Keywords:** GWAS, Neurodevelopment, Cognitive functions, Endophenotype, Parent-of-origin effect

## Introduction

As a species, humans are adept at using communication (both verbal and nonverbal), mental facilities, social interaction abilities and fine motor skills in their everyday lives. These aptitudes mature during neurodevelopment. Some individuals, however, have non-typical neurodevelopment, which is associated with cognitive, motor, behavioral and/or social-cognitive impairments.

\*Correspondence: merete.nordentoft@regionh.dk

<sup>1</sup> CORE - Copenhagen Research Centre for Mental Health, Mental Health Centre Copenhagen, Copenhagen University Hospital, Copenhagen, Denmark

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



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Disorders characterized by these impairment are collectively known as neurodevelopmental disorders, and they often exhibit high comorbidity [1]. Many of these disorders have a strong genetic component, but they often exhibit both genetic and clinical heterogeneity [1–4]. Such a high degree of heterogeneity, in turn, encourages studies into the molecular underpinnings of these disorders.

One strategy which has been proposed as a means to tackle this issue in psychiatric genetics is the use of endophenotypes. Endophenotypes are heritable traits that are (typically) convenient to measure and exhibit an association with the psychiatric condition; more formally, they are said to be heritable traits that are associated with the disease in the population, are primarily disease-state-independent and co-segregate with the disease in families (an additional criterion for complex diseases is that endophenotypes found in affected family members be found in non-affected family members at a higher rate than in the general population) [5]. Endophenotypes can also be quantitative, in which case they should be “milder” in unaffected relatives of affected individuals and correlated with the severity of the disease, and, if this correlation is not due to disease progression or medication, then it could suggest that the correlation with the disease is by way of disease liability [6]. Many traits that can be measured using standardized tests meet these criteria. Pertinent to this study is the case of heritable quantitative traits, which, in turn, may themselves be composites of different measures. For example, it has long been known that general intelligence is heritable [7]. Although the issue of what the intelligence quotient (IQ) itself measures is debated, as are the assumptions about the models estimating its heritability, the overall evidence from twin studies and other family-based studies suggests that a large proportion of the variation in IQ between individuals is due to additive genetic effects [8, 9]. Specifically, indices from subtests of the Wechsler Intelligence Scale for Children also have moderate to high heritabilities [10]. Moreover, specific tests designed to measure various phenotypic expressions of autism spectrum disorder, namely the “Strange Stories” test, which can identify Theory of Mind impairments, and the Social Responsiveness Scale, which provides a quantitative measure of autistic behavioral traits, have both been shown to have modest (“Strange Stories”) to high (Social Responsiveness Scale) heritabilities [11, 12]. In fact, measures from the Social Responsiveness Scale and from the Wechsler Intelligence Scale for Children have been successfully used as endophenotypes in studies of autism spectrum disorder (ASD) and attention deficit/hyperactivity disorder (ADHD) [13, 14]. Lastly, both motor skill and motor learning are also heritable [15, 16], and motor deficits have been suggested

as an endophenotype for schizophrenia [17]. In psychiatry in general, endophenotypes tend to be electrophysiological e.g., electroencephalogram (EEG), eye tracking or certain reflexes [18] or behavioral e.g. gaze direction towards specific facial features [19]. An example of a relatively highly studied gene → behavioral endophenotype → disease pathway is that of the Calcium Voltage-Gated Channel Subunit Alpha1 C (*CACNA1C*) gene, which is a known susceptibility gene for several psychiatric disorders including schizophrenia [20]. A recent study showed that deletions in that gene in mice led to behaviors associated with psychiatric disorders [21].

Even though many genetic studies of the aforementioned traits and disorders (and of complex traits and diseases, in general) have been conducted, these studies together have not identified enough associations to account for the heritabilities of the investigated traits or diseases, a problem known as “the missing heritability” [22]. As genome-wide association studies (GWAS) become larger, more associations are identified at the conventional genome-wide significance threshold. However, there are other reasons why some associations elude the GWAS design, even as sample sizes grow larger: for example, there may be phenotypic heterogeneity not only across individuals, but also in the sense that different studies may use different definitions for disorders, different ascertainment criteria and/or different assessment tools, and, at times, the studied phenotypes themselves might reflect several overlapping underlying abilities. From the genetic perspective, an important reason is that the common GWAS study design, i.e., using only unrelated individuals and modeling only specific types of effects, might not capture all the aspects of the genetic architecture of a trait [22, 23]. Pertinent to this study is the case of the epigenetic phenomenon (i.e., a heritable phenomenon not caused by changes in the DNA sequence itself) known as parent-of-origin effect (POE), whereby the effect of an allele is dependent on its parental origin. Family-based genetic studies, where both parental DNA and proband DNA are available, are ideal for studying these effects. POEs have been implicated in many studies of complex traits and diseases [24]. Studies have shown that, when these effects do operate but are not modeled, they can be missed in traditional GWAS designs [25, 26]. Moreover, the same allele may have opposite effects when inherited paternally vs. maternally [25, 26].

Genomic imprinting is the epigenetic mechanism considered the primary underlying cause of POEs [27]. Imprinted loci are loci at which the two parental alleles are not functionally equivalent (and one of them may even be silenced completely). One molecular mechanism that could lead to imprinting is methylation (the presence

of a methyl group on the DNA nucleotide). Allele-specific methylation in differentially methylated regions (DMRs), or, in this context, imprinting control regions, can lead to differential gene expression depending on the parental origin of the allele [27]. Modification of histones (basic proteins around which DNA is wound to form nucleosomes, a compact package of DNA which makes it possible for the DNA to fit within the nucleus of the cell) can also result in altered gene expression; protein complexes that modify histones covalently can lead to repression of transcription [28]. POEs may also result from mechanisms other than genomic imprinting, for example, bias in transmission of specific types of genetic, such as trinucleotide expansions variation, depending on the sex of the parent [29]. Several disorders which involve genomic imprinting have strong behavioral and cognitive manifestations. Perhaps the most often-cited examples thereof are Prader-Willi syndrome and Angelman syndrome. The genes involved in both of these disorders map to chromosome 15q11q13, but different genes are involved in the two disorders, and they exhibit opposite POEs (paternal for Prader-Willi syndrome and maternal for Angelman syndrome); similarly, the cognitive and behavioral deficits differ between the two disorders [30]. Most cases of these disorders are caused by a deletion of the parentally expressed DNA, but some cases are the result of imprinting defects, leading to aberrant methylation patterns [31, 32]. In the case of complex neurodevelopmental disorders, some notable examples for which POEs have been reported include specific language impairment [33, 34], dyslexia [35] and autism spectrum disorder [36]. A study of 97 traits in mice, where the parent-of-origin of alleles could be determined, found that most of them exhibited POEs, to which a large component of their heritability was attributable. Moreover, the study showed that non-imprinted loci could also exhibit POEs through interaction with imprinted loci [37]. These examples illustrate the importance of considering POEs in studying behavioral and cognitive phenotypes.

Our study aimed to examine both general genetic association as well as parent-of-origin effects, in a deeply phenotyped family-based cohort, in which families were chosen based on the presence (in at least one parent) or absence (in both parents) of a diagnosis of schizophrenia or bipolar disorder, and in which DNA from parents and children was collected, as well as data on a wide array of quantitative neurocognitive, motor, social-cognitive and social-behavioral traits [38]. In prior studies which used this cohort, several of the investigated traits have been shown to differ significantly between children of parents with no diagnosis of schizophrenia or bipolar disorder and children who had at least one parent with a diagnosis of schizophrenia. These included processing speed and

working memory [39], social responsiveness [40], and motor function [41]. Interestingly, these studies did not find similar differences between children of parents with no diagnosis of schizophrenia or bipolar disorder and children who had at least one parent with a diagnosis of bipolar disorder.

The main goal of our study is thus twofold: (i) to find specific genotype–phenotype associations for the quantitative phenotypes from across the aforementioned domains, and (ii) to model POEs in addition to general association to identify associations that would not be captured in case–control GWAS designs. While we do not set out to show that the investigated traits are endophenotypes for specific disorders [as mentioned earlier, some of them have already been used as endophenotypes in previous studies, and they (or similar traits measured by other tests) have been shown to be heritable]), they are all inherently relevant to neurodevelopment in their own right. Moreover, a recent article examining the history of the use of endophenotypes in psychiatry proposed to expand the definition to include transdiagnostic traits, which are not necessarily associated with only one disorder [42]. In this context, identifying genetic variants influencing neurodevelopmental traits is an important endeavor in its own right. To our knowledge, this is the first study which examined these four neurodevelopmental domains in the same cohort, incorporating both general GWAS models and POE models.

## Materials and methods

### Participants

The sample used in this study is part of the Danish High Risk and Resilience Study—VIA 7 (hereafter the VIA 7 study) [38]. The VIA 7 study recruited children aged 7 and their biological parents. Families were recruited from Danish registries on account of having at least one parent with a diagnosis of either schizophrenia spectrum psychosis or bipolar disorder (“high risk” families) or as control families, in which neither parent had schizophrenia or bipolar disorder; however, these disorder were not investigated directly in this study. Overall, of the 402 children with genetic data included in this study (after quality control), 244 come from high risk families (schizophrenia: 147; bipolar disorder: 97), and 158 come from control families. The sample size varies per marker per analysis, as the number of informative children depends on the availability of trait data, marker (genotype) data, and, in the parent-of-origin analyses, parental genotypes as well. We therefore specify the number of informative children (proband) for all significant results individually. Regarding parental data, only genetic data were used in the association tests. After the quality control described below, there were 261.117 trios, 88.0364 child–mother duos,

24.1713 child-father duos, 17.0366 children, 0.352642 parents, 0.173135 mothers and 0.0495053 fathers (as well as 37.9879 parents without children in the dataset), on average per marker, as counted with PREMIM [43], without taking siblings into account. These numbers add up to ~391 (not counting parents without children in the dataset), which is the number of independent children with genetic data in our sample (11 families included a sibling as well).

### Phenotypic data

We investigated eight traits derived from different tests selected from the comprehensive battery of the VIA 7 study: **MABC** (total score from the Danish version of Movement Assessment Battery for Children (Movement ABC-2), 2nd edition [44]. *N.B.*: the norm sample for the Danish version was from the UK, but it has cross-cultural validity [45]); **WISC Coding** [score (total number correct) from the Coding subtest of the Danish version of the Wechsler Intelligence Scale for Children, 4th edition (WISC-IV) [46]]; **WISC Symbol Search** [score (total number correct) from the Symbol Search subtest of the Danish version of the WISC-IV]; **SSR** (score from the Strange Stories—Revised [47], based on the total number of correct answers to 8 mentalizing questions translated into Danish); **SRS** (T-score from the Danish version of the Social Responsiveness Scale (SRS-2) [48], 2nd edition, completed by the child's teacher); **WISC Arithmetic** [score (total correct responses) from the Arithmetic subtest of the Danish version of the WISC-IV]; **WISC Letter-Number Sequencing** [score (number of correct trials) from the Letter-Number Sequencing subtest of the Danish version of the WISC-IV]; **RIST Index** [index score from the Danish version of the Reynold's Intellectual Screening Test (RIST) [49]].

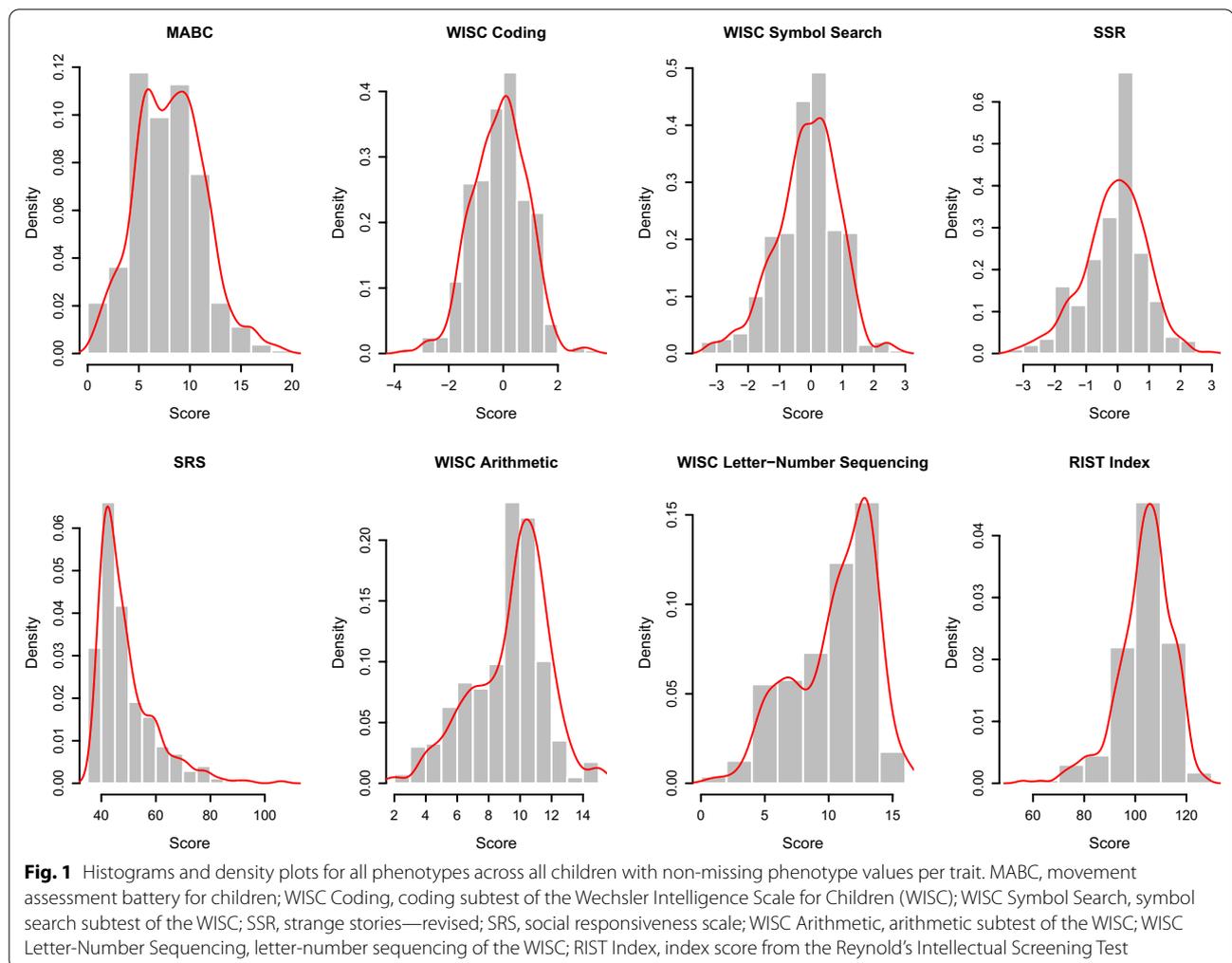
The WISC Arithmetic, WISC Letter-Number Sequencing, RIST Index and MABC scores were age-standardized based on the norms from the manual of each respective test. Where norms were not available for some tests or subtests e.g., when we used the versions of the WISC-IV Coding and Symbol Search subtests for children aged 8 to 16, or when there were no norms (SSR scores), the raw scores were rescaled into Z-scores in SPSS v25.0.0.2 using the mean of the population control subset of VIA 7 children, who were age-matched to the rest of the cohort. The SRS total T-score was not adjusted for age, as this score was not associated with age in children aged 7–15 [12]. More details about these tests can be found in previous publications on the VIA 7 study [39–41]. The distributions of the test scores for each trait are shown in Fig. 1, which contains histograms and density plots for the traits and was generated in R [50] v3.6.3 using the *hist* and *density* functions. We also calculated the pairwise

Pearson's correlation coefficients across the traits in the sample of children with genotypes used in this study, which are shown in Fig. 2. This was done using the Hmisc package v.4.7-0 [51] for R, and the plots were generated with the corrplot package v.0.92 for R [52]. Descriptive statistics for the traits are found in Table 1, which also includes p-values from the Shapiro–Wilk normality test as implemented in the *shapiro.test* function in R.

All traits deviated from normality to some degree. However, as the effective sample for the majority of our tests depends on the parental genotypes, it varies greatly across genetic markers (which are tested individually). Therefore, different subsets of children were used for different markers, and it is not practical to try to transform the scores so as to force them to have a normal distribution, when each transformation will not necessarily work for more than one marker. Moreover, transforming scores in this way would hinder the interpretation of the results, as the spaces between scores would have been changed unevenly, and, therefore, the interpretation of the effect sizes would be problematic. We discuss this at length in a previous paper, where we also examined the difference normalization had made for our top result [53]. Lastly, as we explain below, we used variance components (which assume normality) to correct for relatedness among the children within a given family. As we had only 11 families with more than one child, we investigated the effect of removing a child from each family and not modeling the variance components, and we saw that it had very little impact on our top result [53]; we therefore employ the phenotypic scores as detailed above, without an additional transformation.

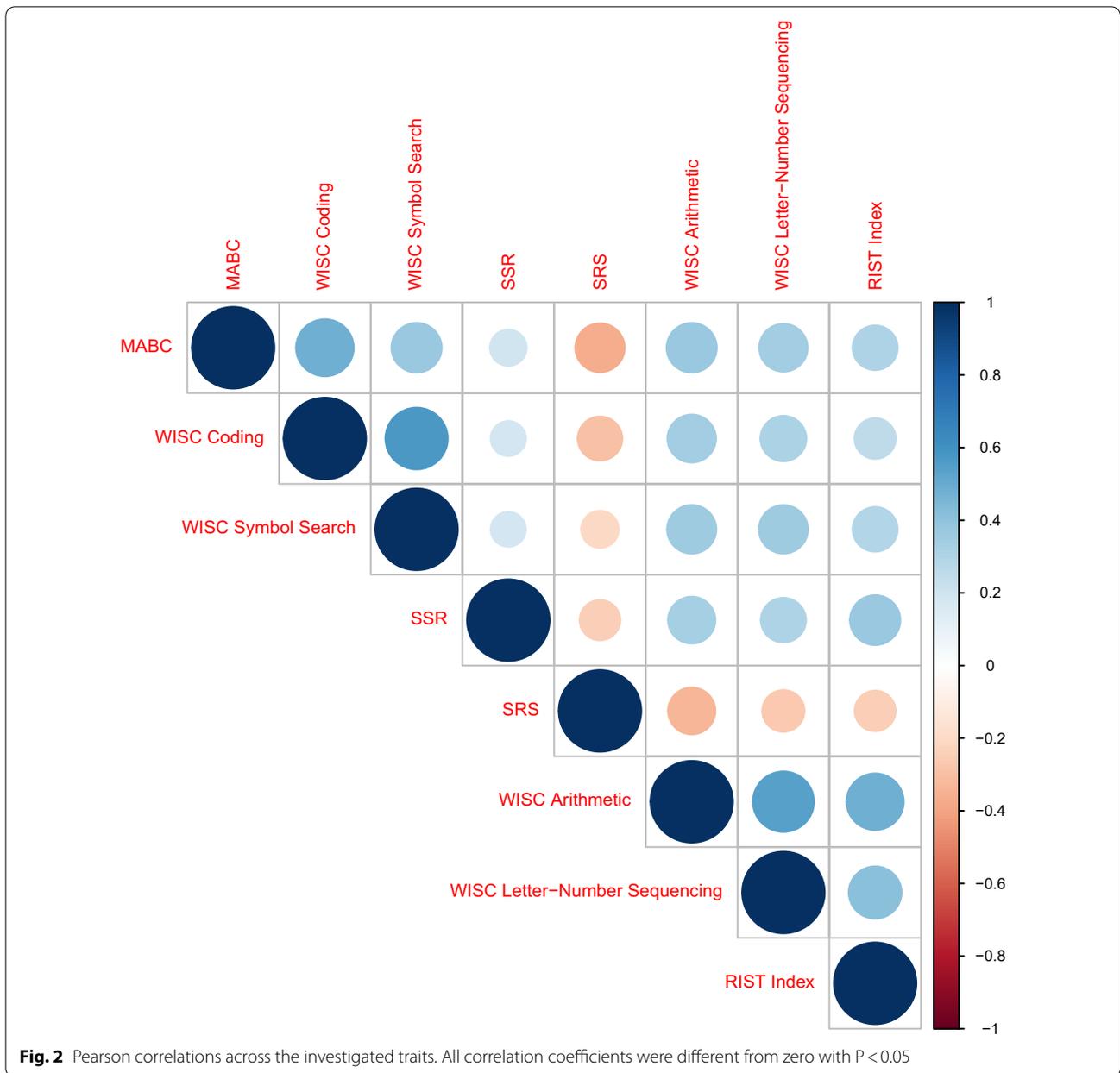
### Genetic data

We had DNA samples from a subset of the VIA 7 study sample, and these were genotyped on the Illumina PsychChip v1-1\_15073391\_C, which had a loci count of 603,144 (according to the information lines in the Illumina manifest file for this array). The dataset has been described in detail in our previous studies [53, 54]. Briefly, the quality control (QC) steps for the samples and markers were as follows: initial QC on raw genetic data: individuals with low call rates or discordant sex information were removed in the first step, as were markers with a Gentrain score < 0.3. At this point 18 individuals had been removed (including one possible duplicate sample), and there were 600,282 markers left in the dataset. Subsequent QC was done with PLINK [55] v.1.90b5.2: individuals and markers with > 1% Mendelian errors were removed (N = 10). Genotypes with remaining Mendelian errors below this threshold were set to missing. Markers with > 5% missing data were removed (at this point all remaining individuals



had <5% missing data). Individuals with extreme heterozygosity rates (with a threshold of  $\pm 3$  SD from the sample mean) were removed ( $N=21$ ). Genetic ancestry was estimated in a principal component analysis (PCA). The threshold for the exclusion of samples was 2 SD above or below the VIA 7 mean for either PC1 or PC2, using the VIA 7 samples and the CEU, CHB, JPT and YRI HapMap samples to create the PC space, as described in a published QC protocol [56]. To reduce bias from population stratification, individuals of divergent ancestry were removed along with their relatives ( $N=36$ ), while the rest of the sample clustered with the CEU individuals. Individuals who exhibited cryptic relatedness or who were less related to biological family members than expected from pedigree information were removed ( $N=13$ ) (the Pi-hat threshold for the exclusion of individuals expected to be unrelated was 0.185). A Hardy–Weinberg Equilibrium (HWE) p-value threshold of  $1 \times 10^{-6}$  was employed for markers, as well

as a minor allele frequency (MAF) threshold of 1% (in founders). Markers with a significant HWE p-value based on the above threshold or MAF below 1% were excluded. We removed one marker per pair in case of pairs of markers with identical positions included in the PsychChip, either using PLINK *--list-duplicate-vars suppress-first*, if the allele codes matched, otherwise prioritizing markers with rsIDs. The number of individuals removed during these steps was 64, including 3 duplicate samples (note that some samples were flagged in more than one step, in cases in which several checks were performed before the final exclusion of samples, namely, after the Mendelian errors check). In total, 1094 genotyped individuals and 299,604 autosomal markers passed these QC steps, from which we further removed 125 indels, for a total of 299,479 autosomal markers. Only this dataset of genotyped markers was used in this study. Positions in the text and tables are in genome build hg19. A minority of markers on the



array had positions in the Illumina manifest file which differed slightly from the ones in dbSNP for the same major build. For that reason we checked all the probes from the manifest file for our array for the top hits in this study, as for the top hit in our previous study [53]. This was done using the UCSC Genome Browser BLAT tool, and, where a marker in our top hits had an rsID, we checked that the SNP was indeed 1 bp away from the probe. Otherwise we checked that the probe mapped to a position 1 bp away from the position in the manifest file. All probes mapped to a position 1 bp away from

the position of interest (the direction depended on the strand), as expected. One marker among the top hits had an incorrect position in the manifest file, but the probe mapped to the right place.

**Statistical analyses—GWAS stage**

In the GWAS stage, we used QTDT (quantitative transmission-disequilibrium test) [57] v2.6.1 for the statistical genetic analyses. MERLIN [58] v1.1.2 was used for estimating identity by descent (IBD) scores for each marker to be used by QTDT. Three tests were performed for

**Table 1** Descriptive statistics for the investigated traits in the study sample. POE: parent-of-origin effect

Trait	Domain	The function measured by the test	Descriptive statistics for phenotypes in the full sample							General test <sup>a</sup>		Paternal POE test <sup>b</sup>		Maternal POE test <sup>b</sup>		
			Children with phenotype data	Mean	Median	Standard deviation	Minimum value	Maximum value	Shapiro-Wilk test p-value	Minimum Number of probands	Maximum Number of probands	Minimum Number of probands	Maximum Number of probands	Minimum Number of probands	Maximum Number of probands	
MABC	Motor	A combined score from tests of manual dexterity, aiming and catching, and balance	399	8.105 8	3.31	3.31	1	19	337	399	196	382	196	382	196	382
WISC Coding	Neuro-cognitive	Processing speed	401	- 0.134 – 0.059	0.992	0.992	- 3.61	3.219	0.0417	338	401	197	384	197	384	384
WISC Symbol Search	Neuro-cognitive	Processing speed	398	- 0.121 – 0.073	0.994	0.994	- 3.292	2.767	8.74 × 10 <sup>-5</sup>	335	398	196	381	196	381	381
SSR	Social-cognitive	Theory of mind	400	- 0.1	0.046	0.999	- 3.227	2.909	1.69 × 10 <sup>-6</sup>	337	400	197	383	197	383	383
SRS	Social-behavioral	Social responsiveness	345	49.014 46	10.441	10.441	37	106	4.68 × 10 <sup>-19</sup>	292	345	172	331	172	331	331
WISC Arithmetic	Neuro-cognitive	Verbal working memory	398	9.475 10	2.396	2.396	2	15	1.18 × 10 <sup>-10</sup>	335	398	196	381	196	381	381
WISC Letter-Number Sequencing	Neuro-cognitive	Verbal working memory	398	10.508 11	3.103	3.103	1	16	7.34 × 10 <sup>-13</sup>	336	398	196	381	196	381	381
RIST Index	Neuro-cognitive	A combined score from tests of non-verbal intelligence (approximation of fluid intelligence) and verbal intelligence (a measure of crystallized intelligence)	401	103.85 105	10.458	10.458	56	127	5.75 × 10 <sup>-10</sup>	338	401	198	384	198	384	384

<sup>a</sup> OTDT chooses informative probands from the pedigree file. Additionally, probands could have been excluded from the tests based on phenotype missingness, genotype missingness and/or problems with their IBD estimation with MERLIN

<sup>b</sup> In addition to the reasons for exclusion in the general test, probands could have also been excluded based on parental genotypes (see "Materials and methods")

each trait-marker combination: a general (i.e., not a POE test) total test of association using all family data (*qtdt -at*), a paternal parent-of-origin total test of association, in which only paternally inherited alleles were used (*-at -op*) and a maternal parent-of-origin total test of association, in which only maternally inherited alleles were used (*-at -om*). The total association model (as opposed to the orthogonal model, which QTDT can also run) is not a TDT, and it was used because it is more powerful in the absence of population stratification [59]. In this model, a combined between/within family component  $X$ , or  $X_{pat}$  and  $X_{mat}$  in the paternal and maternal tests, respectively, denoting the between/within effect on the means, is tested.  $X$  is the effect size reported for the QTDT analyses in this paper.  $X$  is estimated from the data in the full model and is fixed to zero in the null model. The likelihoods of these two models are then assessed through a likelihood ratio test, resulting in a  $\chi^2$  statistic, which can be used to compute a one-sided p-value from the  $\chi^2$  distribution. The tests in this study had one additional free parameter in the full model as compared with the null model, and so the  $\chi^2$  statistic was evaluated with 1 degree of freedom. We included variance components in both models (*-wega*), incorporating an environmental component, a polygenic component and an additive major locus component. This allowed for the use of families with multiple children, although only 11 families included a sibling. Age was taken into account in the scoring of the phenotypes, as explained earlier. For all traits, a covariate for sex was added to both the null model and the full model.<sup>1</sup> The Manhattan plots and the QQ plots were generated with the “qqman” R scripts by Stephen Turner and Daniel Capurso (with the (major update) version from April 19, 2011 for the former type of plot and the version from June 10, 2013 for the latter, available from: <https://github.com/stephenturner/qqman/blob/v0.0.0/qqman.r>). Regional association plots were generated with LocusZoom [60], after converting marker IDs to rsIDs (where possible) using a key from the Illumina website. The QTDT output files were tabulated using an in-house program (included in the Additional file 1), but the statistics for the top hits in our study were also examined manually using the raw QTDT output, and they matched the output of the program.

### Statistical analyses—post hoc tests for paternal and maternal allelic transmission differences

When a POE is detected with one parent, it does not mean that the other parent’s transmissions are

significantly different. It could be that a child effect is significant and appears as such also when looking at paternally inherited alleles or only at maternally inherited alleles separately. Therefore, it is necessary to test for a difference between these parental allele transmissions. This can be achieved by controlling for risk parameters other than the POE parameter by including them in both the null and the full models. QTDT does not allow a free choice of parameters in the null and the full models, but it incorporates a test for the difference in the effects between the paternal and maternal allelic transmissions (*qtdt -at -ot*). In this test, the null model has  $X$ , and the full model has both  $X$  and  $X_{mat}$ . It is not possible to include  $X_{pat}$  instead of  $X_{mat}$  in the full model (to test for a POE when a paternal POE is suspected); therefore, as a precaution, we tested both parameterizations for a known POE using a different program, EMIM [43], which allows to model both parental risk parameters (one at a time in this case) in addition to the child risk parameter, and saw that the overall likelihood of the full model was roughly the same in both cases.<sup>2</sup> Thus, we used this test to filter out associations that are significant in the GWASs when testing paternally inherited alleles and maternally inherited alleles separately, but which do not show a significant difference from the other parent’s transmissions. Note, however, that these models do not test for the simple parental effect at the locus of interest or for the type of POE (if the POE is real), and, for

<sup>2</sup> Consider a previously reported association [34, 54] with a paternal POE for specific language impairment: rs4280164 allele A having a paternal imprinting parameter  $I_p$  (a multiplicative factor by which the probability of disease is multiplied if the child receives a (paternal) copy of the effect allele from their father) = 0.255 with  $P = 2.918 \times 10^{-8}$  (1 degree of freedom). In this case, the null model has all risk parameters fixed to 1, and in the full model  $I_p$  is freely estimated; the two models are assessed using a likelihood ratio test. As the EMIM software (43) (used in the original study) allows many parameterizations, it is also possible to include a free  $R_1$  parameter (the factor by which the disease risk is multiplied if the child has a single copy of the effect allele, and assuming that the risk from the child’s having two risk alleles is  $R_2 = R_1^2$ ) in both the null and full models and a free  $I_p$  parameter only in the full model. This results in a  $\chi^2$  of 28.752840302,  $P = 8.22 \times 10^{-8}$  (1 degree of freedom), suggesting that the paternal POE is still significant even when allowing for a child effect. Importantly, even when estimating the maternal imprinting parameter,  $I_m$  (a multiplicative factor by which the probability of disease is multiplied if the child receives a (maternal) copy of the effect allele from their mother), instead of  $I_p$ , at the same locus (with  $R_1$ ), the likelihood of the full model here provides very similar evidence of association ( $\chi^2$  of 28.752840295436), although the parameter estimates are different. Similarly, even if we do not make the assumption that  $R_2 = R_1^2$  (i.e., we freely estimate both  $R_1$  and  $R_2$  in the null and full models), we still obtain very similar likelihoods, resulting in  $\chi^2 = 25.88917562832$  when estimating  $I_p$  in the full model, and  $\chi^2 = 25.889147279256$  when estimating  $I_m$  in the full model (in both cases including free  $R_1$  and  $R_2$  in both the null and full models), both obtaining  $P = 3.62 \times 10^{-7}$  (1 degree of freedom). We believe that QTDT is doing something similar with its test of allelic difference, and that a test of either combination can be used to determine the presence of a POE, even though QTDT always models the maternal POE effect parameter as the additional one.

<sup>1</sup> *N.B.*: The Danish manual for the SRS we used did not include a sex-adjustment for the T-score; hence, we used a covariate for sex for this trait as well.

that reason, we only use them to test for the presence of a POE and not for effect estimation; we always report the effect of the allele from a test in which only X<sub>mat</sub> or X<sub>pat</sub> are included without X in the full model. Lastly, it should be noted that a POE may be significant with both paternal and maternal transmissions separately and there may still be a significant difference between them, if the association trends are discordant across both parental transmissions (i.e., the same allele may increase the score significantly when inherited from the mother but decrease it significantly when inherited from the father, or vice versa).

**The power and effective sample size of a QTDT analysis**

The power of a QTDT analysis depends on several factors, including: the marker allele frequencies, the effect size, the linkage disequilibrium between the marker and the quantitative trait locus, the number of child genotypes in the analysis and the parental genotypes. Studies which evaluated family-based association methods used simulations of models with the above parameters to estimate the power of those methods. For dichotomous traits, for example, 300 case-mother duos offered reasonable power for detection of child genetic effects [61], when the effects were  $R_1 = 1.5$  and  $R_2 = 2.25$  (see second footnote for an explanation of the parameters), the baseline risk was 0.1, the significance level was 0.05, and the risk allele frequency was 0.3. When strong POEs operate and are included in the model, some methods achieved power of ~90% with as few as 100 case-parents trios [62], with  $I_p = 2.5$  or  $I_m = 2.5$ , a significance level of 0.05, a baseline risk of 0.05 or 0.01 and a risk allele frequency of 0.3 or 0.1, for 20% and 80% of the population, respectively. With regards to quantitative traits, as relevant to this study, we considered published reports of simulations estimating the power of various QTDT models. For example, in the original QTDT paper, assuming a maximum  $D'$ ,  $h^2$  of 0.1, a risk allele frequency of 0.5, a significance level of 0.001 and including parental genotypes, a sample of 480 children (families with a sibship of 1 and parental genotypes available) resulted in a power estimate of 97.4% [57]. In another study, a power of 74% was achieved with a sample size of 200,  $h^2$  of 0.1, and a risk allele frequency of 0.3 [63]. We can translate the effects of an allele into proportion of variance explained (PVE) using the following formula, taken from the supplementary note of a previous study [64]:

$$PVE = \left[ 2 \times \beta^2 \times MAF \times (1 - MAF) \right] / \left[ 2 \times \beta^2 \times MAF \times (1 - MAF) + SE(\beta)^2 \times 2 \times N \times MAF \times (1 - MAF) \right],$$

where  $\beta$  is the effect size,  $SE(\beta)$  is the standard error of  $\beta$ , MAF is the minor allele frequency of the marker (we used

the MAF in founders), and N is the sample size (we used the number of probands). We do this for the top results of our analyses. One further point needs to be taken into consideration with regards to the current set of analyses: the power estimates from the literature are for the QTDT orthogonal model. In the absence of population stratification (as is the case in our study), the total association model can be used, and, all other things being equal, this model has greater power than the orthogonal model [59].

Regarding the sample sizes in the various tests, for the general test, all children who had non-missing genotypes and IBD information for a given marker and non-missing phenotypes were used in the test for that marker. In the POE tests, two groups of children are included: (i) children whose both parents are genotyped and where one parent is homozygous, or whose mother and father have different genotypes (in addition, when paternal parent-of-origin effects are tested, the father must be heterozygous and, when maternal effects are tested, the mother must be heterozygous), and (ii) all children with at least one homozygous parent, even if the other parent has a missing genotype [65]. This may reduce the sample size based on parental genotypes, which is why we report the number of informative probands (probands who meet all the above criteria (for the general test, and, where applicable, the additional criteria for the POE tests) for each association in the top results.

**Statistical analyses—correction for multiple testing and quality measures for GWAS results**

We employed the following strategy for correction for multiple testing in this study: in the GWASs, we present all the associations that met the following two criteria: (i) they pass the conventional genome-wide significance threshold ( $P \leq 5 \times 10^{-8}$ ), (ii) for POE associations, they have  $P \leq 0.0008$  in the test of difference between paternal and maternal alleles, which was calculated as the conventional threshold (0.05) Bonferroni-corrected for the number of post hoc tests for POE associations which met the first criterion ( $n = 63$ ). We then prioritize associations that, in addition to meeting the above two criteria, also meet the following criteria: (iii) they have a p-value (in the GWAS) equal to or below the conventional significance threshold (0.05) Bonferroni-corrected for the actual number of tests performed across all GWASs ( $n = 299,479 \times 24$ ), i.e.,  $P \leq 7 \times 10^{-9}$ ; and (iv) at least 30 children had the minor allele for

the associated marker (*N.B.*: this is not the same as the number of informative probands for QTDT, but rather

it means that at least 30 children in the sample had the minor allele for the marker in question; while this does not guarantee that a specific number of children in a given test had the allele (as this also depended on the factors explained earlier), it could highlight associations for which the effect size is less likely to be biased due to one of the alleles being relatively rare). Associations surviving all four criteria are discussed in more detail in the Results section. For these associations, we also repeated the relevant association test while adding a covariate for the high risk status (HRS) of the family (that is, a dummy variable (0/1) for whether the child is from a family with a parent with schizophrenia or bipolar disorder, or from a family in which neither parent has either of these diagnosis). Additionally, we used EMIM v3.22 [43], a program for multinomial family-based genetic association models, to test for association between the top results (Table 2) and the HRS as a binary outcome. We used a model for child trend analysis [61] in which the factor by which the risk of disease is multiplied when the child has two risk alleles is constrained to be the square of the risk from having one risk allele, or, using the aforementioned notation,  $R_2 = R_1^2$ . In this analyses we used both case and control family subsets, but we did not use controls without parents, since EMIM does not distinguish between controls and individuals with an unknown disease status (which means that parents, who by definition have an unknown HRS, might be used as controls if the child does not have genetic data for a given marker). The p-values for this test are derived from the  $\chi^2$  distribution with one degree of freedom (since only one risk parameter was freely estimated in the full model), and the test statistic comes from twice the difference in the log-likelihoods of a null model (in which the multiplicative risk parameter is fixed to 1) and a full model, in which it is estimated from the data.

QTDT does not output standard errors (SEs) for the estimates it computes. In order to obtain SEs for the observed effect in the top associations in our results we used the following approach: using the  $\chi^2$  statistics from the QTDT output, we calculated the error as  $SE = \sqrt{(X^2/\chi^2)}$ , where  $X$  is the effect size from QTDT. This is an approximation of the SE, because it is calculated from a Wald statistic, whereas QTDT uses a likelihood ratio test for two nested models which differ by the presence of the effect of the genetic variant, but these two methods are at least asymptotically equivalent [66]. Lastly, the genomic inflation factor was calculated for each GWAS in R using the  $\chi^2$  statistics from the QTDT output directly (as QTDT rounds the p-values themselves in the output) as follows: the median of

the observed  $\chi^2$  distribution from each GWAS divided by  $qchisq(0.5, 1)$ .

### Functional annotation of variants and genes

For functional annotation of variants, we used the eQTL-Gen [67] portal and the GTEx V8 portal [68] for finding expression quantitative trait locus (eQTL) associations and PhenoScanner [69] for finding DNA methylation and histone modification associations for the associations meeting our four study-wide criteria for significance. For gene-level annotation we used VarElect [70], which ranks genes based on their association with free text keywords using the GeneCards [71] database.

### Results

Across all 24 GWASs, 88 associations achieved genome-wide significance ( $P \leq 5 \times 10^{-8}$ ), of which 25 were highlighted in the general test and the rest were highlighted in the POE tests. Additional file 2: Fig. S1 shows Manhattan plots for all 24 GWASs, and Additional file 3: Fig. S2 shows the corresponding QQ plots. Across all analyses, the genomic inflation factor ranged from 0.967 to 1.077 (with a mean value of 1.008 and a standard deviation of 0.024). Of the POE associations among the aforementioned 88 associations, only 23 were significant in the test of difference between paternal and maternal alleles after correction for multiple testing (Methods), and the rest were therefore excluded from downstream analyses. The 48 remaining associations are shown in Table 2.

Of the 48 associations that were genome-wide significant and, where applicable, showed a significant difference between paternal and maternal alleles, only 3 met our extra conditions pertaining to the study-wide significance level and a minimum number of 30 probands with the minor allele. Regional association plots for these 3 markers are shown in Fig. 3. We employed these extra criteria to identify more robust associations, especially because very rare alleles could lead to biased effect sizes. Of the 3 associations meeting all four criteria, 2 were with the WISC Arithmetic score and were highlighted in the general test and the remaining association was with the SRS score and showed a maternal POE. Two of these associations were with intragenic variants: rs214831 (general test, associated with WISC Arithmetic) in Transglutaminase 3 gene (*TGM3*) and rs7604835 (maternal POE test, associated with SRS) in the Calcium Channel, Voltage-Dependent, Beta 4 Subunit gene (*CACNB4*). Marker rs214831 was strongly associated with the expression of the gene it was located in, namely, *TGM3*, on eQTLGen ( $P = 5.72 \times 10^{-34}$ ), whereby the A allele was associated with higher expression of the gene; in our study, the effect allele (G) was associated with a lower test score, suggesting that lower expression would

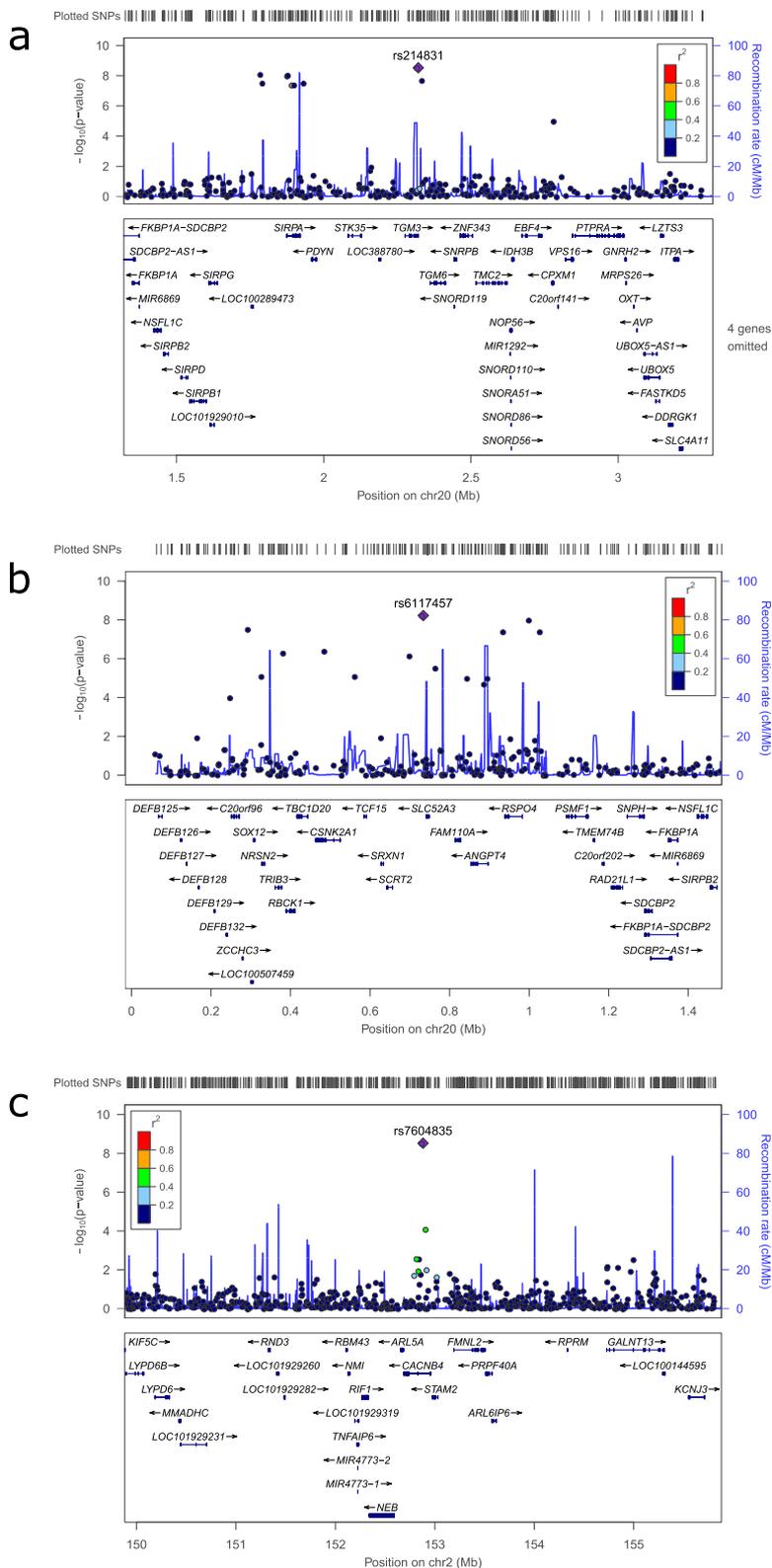
**Table 2** Top results from the GWASs

Test	Trait	Marker ID	Chromosome	Position	MAF	Gene (protein coding)	Effect allele	Other allele	Probands	Effect	SE	$\chi^2$	p-value
General	WISC Arithmetic	psy_rs7826548	8	2955158	0.4717	C5MD1	C	T	389	-0.199	0.036	31.12	$2 \times 10^{-8}$
General	WISC Arithmetic	rs2554728	8	3842999	0.2731	C5MD1	T	C	396	-0.155	0.028	30.47	$3 \times 10^{-8}$
General	WISC Arithmetic	rs11995240	8	3848025	0.1691	C5MD1	T	C	396	-0.016	0.003	29.84	$5 \times 10^{-8}$
General	WISC Arithmetic	rs17068473	8	3854987	0.2854	C5MD1	G	T	396	0.183	0.033	30.72	$3 \times 10^{-8}$
General	WISC Arithmetic	rs4875262	8	3876265	0.3858	C5MD1	G	A	396	0.262	0.046	31.95	$2 \times 10^{-8}$
General	WISC Arithmetic	rs2740929	8	3879918	0.4877	C5MD1	C	T	396	-0.191	0.034	31.03	$3 \times 10^{-8}$
General	WISC Arithmetic	rs2740878	8	3918118	0.3288	C5MD1	A	G	396	-0.309	0.054	33.05	$9 \times 10^{-9}$
General	WISC Arithmetic	rs2552166	8	3920395	0.1734	C5MD1	G	A	396	-0.326	0.057	32.16	$1 \times 10^{-8}$
General	WISC Arithmetic	rs6049002	20	296172	0.3721		C	T	398	0.132	0.024	30.56	$3 \times 10^{-8}$
General	<b>WISC Arithmetic</b>	<b>rs6117457</b>	<b>20</b>	<b>733963</b>	<b>0.2883</b>		<b>T</b>	<b>C</b>	<b>398</b>	<b>-0.359</b>	<b>0.062</b>	<b>33.9</b>	<b><math>6 \times 10^{-9}</math></b>
General	WISC Arithmetic	psy_rs79359757	20	937853	0.07381		C	T	398	-0.053	0.01	30.01	$4 \times 10^{-8}$
General	WISC Arithmetic	rs7261002	20	1002656	0.1835		G	A	398	0.366	0.064	32.9	$1 \times 10^{-8}$
General	WISC Arithmetic	rs6118727	20	1030235	0.4465		A	G	398	0.071	0.013	30.15	$4 \times 10^{-8}$
General	WISC Arithmetic	rs200896	20	1789409	0.4292		A	C	398	-0.326	0.056	33.38	$8 \times 10^{-9}$
General	WISC Arithmetic	psy_rs200888	20	1796461	0.3931	SIRPA	G	T	396	0.138	0.025	30.49	$3 \times 10^{-8}$
General	WISC Arithmetic	psy_rs4813309	20	1880550	0.1331	SIRPA	C	T	397	-0.444	0.078	32.7	$1 \times 10^{-8}$
General	WISC Arithmetic	psy_rs6035018	20	1882954	0.19	SIRPA	T	C	398	-0.392	0.068	33.07	$9 \times 10^{-9}$
General	WISC Arithmetic	exm1519370	20	1896100	0.3994	SIRPA	C	T	388	-0.128	0.023	30.41	$4 \times 10^{-8}$
General	WISC Arithmetic	psy_rs73069290	20	1904515	0.06936	SIRPA	G	T	398	0.068	0.012	30.03	$4 \times 10^{-8}$
General	WISC Arithmetic	rs6035139	20	1936275	0.06014		G	A	398	-0.905	0.163	30.94	$3 \times 10^{-8}$
General	<b>WISC Arithmetic</b>	<b>rs214831</b>	<b>20</b>	<b>2321363</b>	<b>0.3967</b>	<b>TGM3</b>	<b>G</b>	<b>A</b>	<b>398</b>	<b>-0.395</b>	<b>0.067</b>	<b>35.28</b>	<b><math>3 \times 10^{-9}</math></b>
General	WISC Arithmetic	rs6137776	20	2338454	0.2392		T	C	398	0.281	0.05	31.82	$2 \times 10^{-8}$
General	WISC Symbol Search	rs2673776	6	152522812	0.4429	SYNE1	C	A	396	0.163	0.029	31.84	$2 \times 10^{-8}$
General	WISC Symbol Search	exm2270431	6	152644111	0.4565	SYNE1	C	T	396	0.146	0.026	31.44	$2 \times 10^{-8}$
General	WISC Symbol Search	psy_rs9478324	6	152677815	0.05058	SYNE1	G	A	396	0.287	0.052	30.51	$3 \times 10^{-8}$
Maternal	SRS	rs1451197	2	2741446	0.4993		G	A	213	-4.888	0.872	31.42	$2 \times 10^{-8}$
Maternal	SRS	rs2176347	2	45968233	0.4473	PRKCE	T	G	231	-3.441	0.624	30.39	$4 \times 10^{-8}$
Maternal	<b>SRS</b>	<b>rs7604835</b>	<b>2</b>	<b>152881908</b>	<b>0.1315</b>	<b>CACNB4</b>	<b>G</b>	<b>A</b>	<b>287</b>	<b>-1.482</b>	<b>1.927</b>	<b>35.51</b>	<b><math>3 \times 10^{-9}</math></b>
Maternal	SRS	psy_rs77672109	4	35576468	0.01662		C	A	304	-58.859	10.144	33.67	$7 \times 10^{-9}$
Maternal	SRS	rs11934637	4	92981939	0.0289		T	G	304	-34.434	5.853	34.61	$4 \times 10^{-9}$
Maternal	SRS	psy_rs146416593	5	60662093	0.01592	ZSWIM6	G	A	324	-33.216	5.844	32.31	$1 \times 10^{-8}$
Maternal	SRS	psy_rs79166730	5	175229558	0.01951	CPLX2	G	T	321	-37.303	5.817	41.13	$1 \times 10^{-10}$

**Table 2** (continued)

Test	Trait	Marker ID	Chromosome	Position	MAF	Gene (protein coding)	Effect allele	Other allele	Probands	Effect	SE	$\chi^2$	p-value
Maternal	SRS	exm573219	6	116325108	0.01879	<i>FRK</i>	G	A	325	-58.949	9.981	34.88	$4 \times 10^{-9}$
Maternal	SRS	psy_rs78989171	8	49645477	0.01301	<i>EFCAB1</i>	G	T	308	-58.867	10.204	33.28	$8 \times 10^{-9}$
Maternal	SRS	exm754630	9	74319677	0.01806	<i>TMEM2</i>	T	C	322	-26.113	4.085	40.87	$2 \times 10^{-10}$
Maternal	SRS	rs1681993	12	63338414	0.01084		G	A	324	-58.96	10.14	33.81	$6 \times 10^{-9}$
Maternal	SRS	rs7956933	12	63345018	0.01085		A	G	323	-58.964	10.151	33.74	$6 \times 10^{-9}$
Maternal	SRS	psy_rs10860381	12	99309750	0.01375	<i>ANKK1B</i>	G	A	305	-58.851	10.206	33.25	$8 \times 10^{-9}$
Maternal	WISC Arithmetic	rs11784069	8	2119582	0.3736		T	G	271	0.341	0.061	31.1	$2 \times 10^{-8}$
Maternal	WISC Arithmetic	rs7261002	20	1002656	0.1835		G	A	314	0.661	0.116	32.68	$1 \times 10^{-8}$
Maternal	WISC Symbol Search	rs2256135	6	152464839	0.4566	<i>SYNE1</i>	G	A	262	0.264	0.047	31.95	$2 \times 10^{-8}$
Paternal	RIST index	exm693219	8	28929739	0.02746	<i>KIF13B</i>	G	A	371	42.012	7.191	34.13	$5 \times 10^{-9}$
Paternal	SRS	exm109001	1	156314440	0.01517	<i>TSACC</i>	T	G	320	-42.257	7.231	34.15	$5 \times 10^{-9}$
Paternal	SRS	psy_rs191695175	8	122841477	0.01016		T	C	326	-47.854	5.665	71.36	$3 \times 10^{-17}$
Paternal	SRS	rs16908233	11	21604897	0.01158		A	G	317	-34.271	5.86	34.2	$5 \times 10^{-9}$
Paternal	SRS	psy_rs117476444	13	113442872	0.02168	<i>ATP11A</i>	G	A	320	-33.185	5.847	32.21	$1 \times 10^{-8}$
Paternal	WISC Arithmetic	rs11784069	8	2119582	0.3736		T	G	271	-0.273	0.049	30.71	$3 \times 10^{-8}$
Paternal	WISC Arithmetic	rs2740939	8	3872513	0.4899	<i>CSMD1</i>	C	A	253	-0.244	0.044	30.73	$3 \times 10^{-8}$

Results meeting criteria (i) and (ii) (see "Materials and methods") are shown. Results meeting criteria (iii) and (iv) are shown in boldface. MAF: minor allele frequency (in founders; does not necessarily correspond to the effect allele frequency); SE, standard error



**Fig. 3** Regional association plots for associations surviving all four statistical quality criteria (Methods). **a** rs6117457 (general test, WISC Arithmetic); **b** rs6117457 (general test, WISC Arithmetic); **c** rs7604835 (maternal POE test, SRS)

be associated with a lower score. It was also associated with the expression of *PTPRA* in the basal ganglia on GTEx ( $P=0.000022$ ), with allele G being associated with lower expression. This marker remained at least nominally significant when adding a covariate for high risk status (i.e. for whether the child comes from a high risk family or a control family) to the model ( $P=0.0312$ ). Marker rs7604835, which showed a maternal POE in our study, was associated with multiple DNA methylation and histone modification sites on PhenoScanner (minimum  $P=1.19 \times 10^{-45}$ ), based on evidence from two different studies [72, 73]. This provides further support for the association with a POE at this locus. This marker remained genome-wide significant when adding a covariate for high risk status ( $P=2 \times 10^{-9}$ ). The last association which met all four criteria, namely, between rs6117457 and WISC Arithmetic in the general test, did not implicate any protein-coding gene, and we could not find any relevant prior association with it in the literature or functional databases. This marker did not remain significant when adding a covariate for high risk status ( $P=0.0765$ ). Translating the effects of the top markers into PVEs, we get: 0.078, 0.08 and 0.11 for rs6117457, rs214831 and rs7604835, respectively. The associations adjusted for HRS were in the same direction as before in all cases. It should be noted, however, that the interpretation of the models with the covariate for HRS can be difficult: both WISC Arithmetic and SRS are associated with the child's schizophrenia family status in the VIA 7 study [39, 40]; since the covariate in this case may imply some genetic predisposition to schizophrenia, a disorder which is genetically correlated with cognitive traits [74], the same SNP could have some association with both the psychiatric disorder and the phenotype of interest. Furthermore, both schizophrenia and bipolar disorder are complex, meaning they have both genetic and environmental risk factors [75, 76]. Thus, the high risk status of the family, determined by the presence of a psychiatric diagnosis in one of the parents, is influenced both by genetic factors and environmental factors; the parental genetic factors influence both the child's genetics (the exposure) and the high risk status of the family (the parent's illness and potential covariate), which could influence the outcome in the child (the investigated trait). Similarly, environmental factors, which may be unmeasured (or external factors in general e.g., parental IQ), could influence both the high risk status of the family and the investigated trait in the child. In this scenario, adjusting for the covariate may reduce bias from possible confounding but introduce collider bias. A further complication would be the fact that most of our tests were for POEs, which limit the genetic causal path but not the causal path of the high risk status on which families were ascertained in this

study. We therefore tested whether these markers were themselves associated with the high risk status as the outcome; none of the markers in Table 2 were associated with it after Bonferroni correction for multiple testing, and the top three markers were not nominally associated even before correction. Thus, if, for these markers, the high risk status of the family is not associated with the genetic exposure, then this eliminates both the potential confounding and potential collider bias from the model, even if HRS is not included as a covariate. Even though the high risk status refers to the parent and not the child's phenotype, this lack of association could suggest that the traits highlighted in Table 2 might not be useful endophenotypes for schizophrenia or bipolar disorder, but might nonetheless be associated with other disorders.

Most of the associated markers in Table 2 (29 out of 48) fall within protein-coding genes. In total, 15 unique genes are implicated by at least one genome-wide significant association [meeting criteria (i) and, where relevant, (ii)] with a variant within them: *ANKS1B*, *ATP11A*, *CACNB4*, *CPLX2*, *CSMD1*, *EFCAB1*, *FRK*, *KIF13B*, *PRKCE*, *SIRPA*, *SYNE1*, *TGM3*, *TMEM2*, *TSACC* and *ZSWIM6*. Given that the associations in Table 2 were with the Arithmetic and Symbol Search subtests of the WISC, the SRS and the RIST, we used the following terms together with the gene names when running the VarElect analysis: autism OR "working memory" OR behavior OR communication OR intelligence OR "processing speed" OR "Wechsler Intelligence Scale for Children" OR "Reynolds Intellectual Screening Test" OR "Social Responsiveness Scale" OR schizophrenia OR "bipolar disorder". The last two terms were added because they represent the disorders based on which the VIA sample had been ascertained. Fourteen out of the fifteen genes were directly associated with at least one of the terms (i.e., the gene's GeneCard contained the term), with the average number of associated terms per gene being  $4.43 (\pm 2.41)$ . Two genes were associated with 8 terms, the highest number of terms any one gene was associated with: Ankyrin Repeat and Sterile Alpha Motif Domain-containing Protein 1B (*ANKS1B*), and Synaptic Nuclear Envelope Protein 1 (*SYNE1*). Additional file 4: Table S1 lists all direct associations between the terms and the genes and a discussion of the scores. The gene with the highest VarElect score was CUB And Sushi Multiple Domains 1 (*CSMD1*), and the gene with the highest average disease causing likelihood was the aforementioned *CACNB4*.

## Discussion

Our study investigated eight neurocognitive, motor and social-cognitive and social-behavioral functions using a family-based GWAS design, including a general association test as well as tests of parent-of-origin effect tests.

We have identified 48 genome-wide significant associations, of which 3 met our study-wide significance threshold. Our results highlighted several protein-coding genes, some of which have been implicated in prior genetic analyses of relevant phenotypes.

Two genes were highlighted through associations which met all four of our significance criteria: *TGM3* and *CACNB4*. The association with *TGM3* was further supported by the marker's being an eQTL for the gene. This gene is involved in terminal epidermal differentiation and has been implicated in some cancers [77, 78]. In our study, the marker in this gene was associated with a measure of working memory. Interestingly, previous studies have found relevant associations between the gene and related phenotypes: a study of the RNA blood transcriptome of patients with Alzheimer's disease (AD), a disease which involves severe memory impairments, found that the largest expression fold change among differentially expressed genes across AD cases and controls was with *TGM3* [79]. Genes of the same family have been implicated in several neurodegenerative diseases [80]. Also of note, the associated marker in our study was also a brain eQTL for *PTPRA*, a gene which is important for hippocampal neuronal migration; mice deficient for the *PTPRA* protein exhibit impairments in learning and short-term memory [81]. The association between social responsiveness (SRS) and *CACNB4* was with a maternal POE. This marker was also associated with methylation and histone modifications sites, providing further support for a POE. The gene encodes a member of the beta subunit family of voltage-dependent calcium channels, and it belongs to a family of genes which has been implicated in several psychiatric and neurodevelopmental disorders, including autism spectrum disorder, across many studies [82]. The subunit encoded by *CACNB4*, specifically, is highly expressed in the brain and is prominent in the cerebellum [83]. A recent study found that a pathogenic missense variant in this gene resulted in a severe neurodevelopmental impairment which included intellectual disability, language impairment, movement impairment and seizures [84]. When adding a covariate for high risk status to the statistical models for the top associations, we observe that it either slightly improved the association (with SRS, maternal POE test) or reduced it drastically (with WISC Arithmetic, general test). Whether or not it is appropriate to include this covariate in the model depends on the causal paths between the genetic variant, the trait, and the covariate, which are complex and not known. Hence, the interpretation of these post hoc tests should be done with caution.

Among the other genes in Table 2, three genes were highlighted in the functional annotation either as having the highest VarElect score or as being associated

with the largest number of terms: *CSMD1*, *ANKS1B* and *SYNE1*. *CSMD1* is of particular interest because it has been implicated in schizophrenia [85–87]. Interestingly, in our study, this gene was implicated through markers associated with a measure of working memory; a study of this gene reported that a schizophrenia risk variant in *CSMD1* was associated with spatial working memory [88]. This could illustrate the effect of a genetic variant on an endophenotype for schizophrenia. In this context it is also important to note a proposal to redefine the notion of endophenotype in psychiatry to allow it to include transdiagnostic traits that may be shared across several disorders [42]. *ANKS1B* was implicated through the association between the marker *psy\_rs10860381* and social responsiveness in the maternal POE test. This gene encodes an activity dependent postsynaptic effector protein highly expressed in the brain, and it has been implicated in a wide array of neurodevelopmental phenotypes [89]. Importantly, haploinsufficiency of this gene in a mouse model resulted in impaired social interaction and sensorimotor dysfunction, which are core features of autism spectrum disorder [90]. Even more importantly, this gene exhibits allelic expression imbalance in the brain, which could be an outcome of genomic imprinting (which could result in a POE), although this is only one possible explanation [89]. *SYNE1* was implicated through the associations between several markers and processing speed (WISC Symbol Search) in the general test. The gene encodes a protein that is involved in anchoring specialized myonuclei underneath neuromuscular junctions, but it is also expressed in the brain—predominantly in the cerebellum [91]. It has been implicated in a recessive form of cerebellar ataxia, which may also include cognitive deficits [91]. Interestingly, individuals with *SYNE1* mutations exhibit processing speed deficits compared with controls [92], which is in line with our result showing association between this gene and processing speed in the general test. Both *SYNE1* and *TGM3* have been highlighted in a study of de novo mutations in autism spectrum disorder [93].

Some of the other associations in Table 2 are also of note. The paternal POE association between social responsiveness and *rs191695175* was the most significant association in our study. The minor allele frequency for this marker was very low at ~0.01 (in founders), which could lead to a biased effect size. However, this marker is found on chromosome 8 in chromosomal band 8q24.13, a locus which was part of a suggestive linkage peak for the same trait, namely, SRS, in a genome-wide linkage study [13]. The same locus also showed linkage to SRS in addition to an anxiety score and a score for pragmatic language skills, in another study [94]. These studies, however, did not

model POEs. Thus, even though we may not be able to trust the estimated effect size for this locus, the association itself might be valid and supported by previous studies, and it is possible that the POE, if it indeed operates at this locus, contributed to the stronger signal in this study as compared with previous studies. There have been other previous studies which included similar phenotypes, such as social interaction and social communication (neither was measured with the SRS), but they did not model POEs, and their significant results do not overlap with ours [95, 96]. We also observe an interesting association trend with marker rs11784069: allele T, when inherited from the mother, is associated with a higher WISC Arithmetic score (better working memory functions), but, when inherited from the father, it is associated with a lower score (Table 2). This is an illustration of the phenomenon mentioned in the introduction, namely, opposite POEs of different parental types at the same locus, which has been observed for other quantitative traits in humans. This marker is a highly significant eQTL for *MYOM2* on eQTLGen ( $P=3.2717 \times 10^{-310}$ ) and GTE<sub>x</sub> ( $P=9.5 \times 10^{-20}$ ) in whole blood. Interestingly, the mouse ortholog of this gene, *Myom2*, was significantly upregulated and had the fifth largest fold change among upregulated genes in the hippocampus of memory-enhanced mice in one study [97], which is relevant for the association in our study, as the WISC Arithmetic score is a measure of working memory.

#### The top results in the context of endophenotypes and the investigated domains

The traits implicated by the top results in our study, namely, SRS (social responsiveness) and WISC Arithmetic (working memory), had been proposed as endophenotypes for ASD and ADHD, respectively [13, 98–100]. However, these studies did not identify links between specific genes and these endophenotypes at a genome-wide significant level; they focused on linkage analyses or candidate genes, and, where association was modeled, it was only suggestive. Thus, our study provides genetic evidence for the missing piece in the pathway from gene to disorder through endophenotype, namely: *TGM3* → working memory → ADHD and *CACNB4* → social responsiveness → ASD, through the top genetic associations we identified. Similarly, memory impairments, including verbal working memory impairment, are common feature of schizophrenia [101], suggesting further pathways between *TGM3*, *CSMD1* and *PTPRA* and schizophrenia through the working memory endophenotype. The highlighted associations in Table 2 belong to the neurocognitive and social-behavioral

domains. This does not mean that traits from the other domains would not make good endophenotypes; our study did not examine that, and the lack of genetic association could result from lower heritability for those traits and/or insufficient sample sizes.

#### Limitations of our study

Our results should be evaluated in the light of several potential limitations. Firstly, our study sample was a family-based sample, and, as such, not a very large one. While this has the advantage of our being able to have a deeply phenotyped sample, it can be detrimental to genetic association studies. While, as shown in previous simulations studies of QTDT models, our sample should be large enough to detect some effects, it is expected that only strong effects could be detected in our sample, which can explain why the majority of our genome-wide significant associations were intragenic. It should also be emphasized that some of the effect sizes could be overestimated due to confounding. As it is difficult to determine the appropriateness of the adjustment for high risk status, it should be borne in mind that the effects for some associations might not be accurate. However, since the GWASs were performed with the goal of discovering new genetic associations for downstream analyses and not for estimating their effects, we adopted this approach rather than potentially over-adjust the models, as discussed earlier. Another limitation is that we did not have a suitable replication sample which included the same phenotypes and genetic data from children and parents. Although our candidate genes have been highlighted in previous studies of related traits, providing more credibility to their association with our traits, the associations with specific variants need to be replicated in an independent sample.

#### Future perspectives

It has been shown that the heritability of cognitive ability increases from childhood to young adulthood [102]. Interestingly, a similar trend (reaching its peak around age of 13 for girls and 14 for boys) was observed for height [103]. When the proportion of phenotypic variance explained by genetics increases, the proportion of the variance explained by the environment decreases, and vice versa. In the case of height, this trend could reflect the effect of early childhood living conditions and/or prenatal environmental factors [103]. For cognitive ability, the authors theorize that this trend could be a result of genotype-environment correlation, whereby their genetics influences children increasingly in selecting, modifying and creating their own experiences as they grow up [102]. From the statistical genetic perspective, a higher

heritability means that more genetic associations could be identified if the sample of children were studied when they are older; this could mean that repeating the analyses within the VIA sample with these functions measured in early adulthood could result in further associations. Furthermore, functional studies of the genes highlighted in our study could provide further insight into the molecular etiologies of the neurodevelopmental disorders whose endophenotypes were investigated in this study.

## Conclusions

Our study identified several candidate genes for social-behavioral and neurocognitive functions, implicated either through a general test, or a test of POEs; associations in the latter test were also supported by external studies which had identified methylation or histone modification sites associated with the relevant marker. Importantly, most of our genome-wide significant associations were within protein-coding genes, and many of these had previously been implicated in studies of related traits and disorders, although many of these previous associations were with rare and/or deleterious mutations. Our study provides further evidence to the effect that common variants may influence related traits in individuals not diagnosed with severe mental disorders, and it further supports a role for the highlighted genes in the studied traits, which can be seen as a replication of those genes' implications in the previous studies. We did not identify significant associations for traits in some of the other functions/domains included in this study; this could be the result of the lower heritability of those traits, as well as potentially smaller effects that could not be discovered in the VIA sample. Our results also illustrate the usefulness of modeling POEs in human genetic studies, and, while previous studies focused on an array of quantitative non-social-cognitive, non-social-behavioral, and non-neurocognitive traits, our study highlights the presence of potential POEs in several of these traits studied in a systematic way, thus providing further evidence for this phenomenon in humans.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12993-022-00198-0>.

**Additional file 1.** Archive file containing std\_qtdt, which can parse and tabulate the QTD output.

**Additional file 2: Figure S1.** Manhattan plots for all 24 GWASs.

**Additional file 3: Figure S2.** QQ plots for all 24 GWASs.

**Additional file 4: Table S1.** Direct associations from VarElect for genes in Table 1.

## Acknowledgements

The authors would like to express their gratitude to the dedicated families participating in the study. We would also like to thank Heather Cordell for very helpful discussions on the model parameterizations of parent-of-origin effects, and Oskar Hougaard Jepsen for a valuable discussion on covariate adjustments and potential biases in the models.

## Author contributions

RN conceived the study, performed the QC of the genetic data, performed the genetic and statistical analyses, analyzed the results, wrote the paper; RZ wrote the program that tabulated the QTD output; JO performed data management for VIA 7, assisted with the QC of the pedigree information, performed the standardization of test scores which lacked norms; CAJC, NH, DE, KSS, BKB, ANG, DLG contributed to the VIA 7 data collection and/or pilot study; CAJC, NH, BKB were responsible for the non-genetic study of the investigated traits in the VIA 7 cohort and provided information about the trait scores relevant to this study; J-BG oversaw sample preparation and genotyping and performed initial QC on the raw genetic data; TW designed and oversaw the genetic part of the VIA 7 study; KJP, AAET, JRMJ, OM, MN contributed to the conception of the VIA 7 study and its design, coordination and funding applications. All authors have read and approved the manuscript.

## Funding

The VIA 7 project is supported by the Mental Health Services of the Capital Region of Denmark (Region Hovedstadens Psykiatri), the Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH), Aarhus University, the Tryg Foundation and the Beatrice Surovell Haskell Fund for Child Mental Health Research of Copenhagen. The funding bodies had no influence on the study design, analysis, or interpretation or on the writing of the manuscript.

## Availability of data and materials

Access to the dataset used in the current study is available from the corresponding author upon reasonable request. The program used to tabulate the QTD output is available in an additional file accompanying this article.

## Declarations

### Ethics approval and consent to participate

The study was approved by the Danish Data Protection Agency and follows all laws concerning the processing of personal data. Permission to draw data from registers was granted by the Danish Ministry of Health. The study protocol was sent to the Danish Committee on Health Research Ethics, who decided that ethical approval was not needed due to the observational nature of the study. The genetic part of the study obtained ethical approval from the outset of the study and The Danish High Risk and Resilience Study –VIA 7 was later incorporated into the protocol (Arv og Miljø—genetics and environment) as an appendix, which has then been approved by the ethics committee [ARV OG MILJØ: betydning for psykisk sygdom hos børn og unge (H-B-2009-026)]. Written informed consent was obtained from all adult participants and from the legal guardians of participating children.

### Consent for publication

Not applicable.

### Competing interests

The authors have no competing interests to declare, but TW states that he has acted as a lecturer and scientific counselor to H. Lundbeck A/S. DE has been employed by H. Lundbeck A/S from March until August 2020.

### Author details

<sup>1</sup>CORE - Copenhagen Research Centre for Mental Health, Mental Health Centre Copenhagen, Copenhagen University Hospital, Copenhagen, Denmark. <sup>2</sup>iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Aarhus, Denmark. <sup>3</sup>Institute of Biological Psychiatry, Mental Health Centre Sct. Hans, Mental Health Services Copenhagen, Roskilde, Denmark. <sup>4</sup>Mental Health Centre for Child and Adolescent Psychiatry - Research unit, Mental Health Services in the Capital Region of Denmark, Copenhagen, Denmark. <sup>5</sup>Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. <sup>6</sup>Psychosis Research Unit, Aarhus University Hospital - Psychiatry, Aarhus, Denmark.

<sup>7</sup>Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark. <sup>8</sup>Division of Child and Adolescent Psychiatry, Department of Psychiatry, Hospital University Lausanne, Lausanne University, Lausanne, Switzerland. <sup>9</sup>Center for Neuropsychiatric Schizophrenia Research and Center for Clinical Intervention and Neuropsychiatric Schizophrenia Research, Mental Health Services in the Capital Region of Denmark, Copenhagen, Denmark.

Received: 24 February 2022 Accepted: 9 September 2022  
Published online: 01 December 2022

## References

- Parenti I, Rabaneda LG, Schoen H, Novarino G. Neurodevelopmental disorders: from genetics to functional pathways. *Trends Neurosci*. 2020;43(8):608–21 (**Epub 2020/06/09**).
- Lenartowicz A, Loo SK. Use of EEG to diagnose ADHD. *Curr Psychiatry Rep*. 2014;16(11):498 (**Epub 2014/09/23**).
- Masi A, DeMayo MM, Glozier N, Guastella AJ. An overview of autism spectrum disorder, heterogeneity and treatment options. *Neurosci Bull*. 2017;33(2):183–93 (**Epub 2017/02/19**).
- Bishop DVM. What causes specific language impairment in children? *Curr Dir Psychol Sci*. 2006;15(5):217–21 (**Epub 2008/11/15**).
- Gottesman II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry*. 2003;160(4):636–45 (**Epub 2003/04/02**).
- Almasy L, Blangero J. Endophenotypes as quantitative risk factors for psychiatric disease: rationale and study design. *Am J Med Genet*. 2001;105(1):42–4 (**Epub 2001/06/27**).
- Blokland GAM, Mesholam-Gately RI, Touloupoulou T, Del Re EC, Lam M, DeLisi LE, et al. Heritability of neuropsychological measures in schizophrenia and nonpsychiatric populations: a systematic review and meta-analysis. *Schizophr Bull*. 2017;43(4):788–800 (**Epub 2016/11/23**).
- Plomin R, von Stumm S. The new genetics of intelligence. *Nat Rev Genet*. 2018;19(3):148–59 (**Epub 2018/01/18**).
- Visscher PM, Hill WG, Wray NR. Heritability in the genomics era—concepts and misconceptions. *Nat Rev Genet*. 2008;9(4):255–66 (**Epub 2008/03/06**).
- van Soelen IL, Brouwer RM, van Leeuwen M, Kahn RS, Hulshoff Pol HE, Boomsma DI. Heritability of verbal and performance intelligence in a pediatric longitudinal sample. *Twin Res Hum Genet*. 2011;14(2):119–28 (**Epub 2011/03/24**).
- Ronald A, Viding E, Happe F, Plomin R. Individual differences in theory of mind ability in middle childhood and links with verbal ability and autistic traits: a twin study. *Soc Neurosci*. 2006;1(3–4):412–25 (**Epub 2008/07/18**).
- Constantino JN, Todd RD. Autistic traits in the general population: a twin study. *Arch Gen Psychiatry*. 2003;60(5):524–30 (**Epub 2003/05/14**).
- Lowe JK, Werling DM, Constantino JN, Cantor RM, Geschwind DH. Social responsiveness, an autism endophenotype: genomewide significant linkage to two regions on chromosome 8. *Am J Psychiatry*. 2015;172(3):266–75 (**Epub 2015/03/03**).
- Pineda DA, Lopera F, Puerta IC, Trujillo-Orrego N, Aguirre-Acevedo DC, Hincapie-Henao L, et al. Potential cognitive endophenotypes in multi-generational families: segregating ADHD from a genetic isolate. *Atten Deficit Hyp Disord*. 2011;3(3):291–9 (**Epub 2011/07/23**).
- Missitzi J, Gentner R, Misitzi A, Geladas N, Politis P, Klissouras V, et al. Heritability of motor control and motor learning. *Physiol Rep*. 2013;1(7):e00188 (**Epub 2014/04/20**).
- Williams LR, Gross JB. Heritability of motor skill. *Acta Genet Med Gemellol*. 1980;29(2):127–36 (**Epub 1980/01/01**).
- Burton BK, Hjorthoj C, Jepsen JR, Thorup A, Nordentoft M, Plessen KJ. Research review: do motor deficits during development represent an endophenotype for schizophrenia? A meta-analysis. *J Child Psychol Psychiatry*. 2016;57(4):446–56 (**Epub 2015/11/19**).
- Iacono WG, Vaidyanathan U, Vrieze SJ, Malone SM. Knowns and unknowns for psychophysiological endophenotypes: integration and response to commentaries. *Psychophysiology*. 2014;51(12):1339–47 (**Epub 2014/11/13**).
- Constantino JN, Kennon-McGill S, Weichselbaum C, Marrus N, Haider A, Glowinski AL, et al. Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature*. 2017;547(7663):340–4 (**Epub 2017/07/13**).
- Moon AL, Haan N, Wilkinson LS, Thomas KL, Hall J. CACNA1C: association with psychiatric disorders, behavior, and neurogenesis. *Schizophr Bull*. 2018;44(5):958–65 (**Epub 2018/07/10**).
- Dedic N, Pohlmann ML, Richter JS, Mehta D, Czamara D, Metzger MW, et al. Cross-disorder risk gene CACNA1C differentially modulates susceptibility to psychiatric disorders during development and adulthood. *Mol Psychiatry*. 2018;23(3):533–43 (**Epub 2017/07/12**).
- Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet*. 2010;11(6):446–50 (**Epub 2010/05/19**).
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461(7265):747–53 (**Epub 2009/10/09**).
- Kong A, Steinthorsdottir V, Masson G, Thorleifsson G, Sulem P, Besenbacher S, et al. Parental origin of sequence variants associated with complex diseases. *Nature*. 2009;462(7275):868–74 (**Epub 2009/12/18**).
- Benonisdottir S, Oddsson A, Helgason A, Kristjansson RP, Sveinbjornsson G, Oskarsdottir A, et al. Epigenetic and genetic components of height regulation. *Nat Commun*. 2016;7:13490 (**Epub 2016/11/17**).
- Mozaffari SV, DeCaro JM, Shah SJ, Sidore C, Fiorillo E, Cucca F, et al. Parent-of-origin effects on quantitative phenotypes in a large Hutterite pedigree. *Communications biology*. 2019;2:28 (**Epub 2019/01/25**).
- Lawson HA, Cheverud JM, Wolf JB. Genomic imprinting and parent-of-origin effects on complex traits. *Nat Rev Genet*. 2013;14(9):609–17 (**Epub 2013/08/07**).
- Weaver JR, Bartolomei MS. Chromatin regulators of genomic imprinting. *Biochem Biophys Acta*. 2014;1839(3):169–77 (**Epub 2013/12/19**).
- Pearson CE. Slipping while sleeping? Trinucleotide repeat expansions in germ cells. *Trends Mol Med*. 2003;9(11):490–5 (**Epub 2003/11/08**).
- Cassidy SB, Dykens E, Williams CA. Prader-Willi and Angelman syndromes: sister imprinted disorders. *Am J Med Genet*. 2000;97(2):136–46 (**Epub 2001/02/17**).
- Buiting K. Prader-Willi syndrome and Angelman syndrome. *Am J Med Genet C Semin Med Genet*. 2010;154C(3):365–76 (**Epub 2010/08/31**).
- Chamberlain SJ, Lalande M. Neurodevelopmental disorders involving genomic imprinting at human chromosome 15q11–q13. *Neurobiol Dis*. 2010;39(1):13–20 (**Epub 2010/03/23**).
- Nudel R, Simpson NH, Baird G, O'Hare A, Conti-Ramsden G, Bolton PF, et al. Associations of HLA alleles with specific language impairment. *J Neurodevel Disord*. 2014;6(1):1 (**Epub 2014/01/18**).
- Nudel R, Simpson NH, Baird G, O'Hare A, Conti-Ramsden G, Bolton PF, et al. Genome-wide association analyses of child genotype effects and parent-of-origin effects in specific language impairment. *Genes Brain Behav*. 2014;13(4):418–29 (**Epub 2014/02/28**).
- Pettigrew KA, Frinton E, Nudel R, Chan MTM, Thompson P, Hayiou-Thomas ME, et al. Further evidence for a parent-of-origin effect at the NOP9 locus on language-related phenotypes. *J Neurodev Disord*. 2016;8:24 (**Epub 2016/06/17**).
- Connolly S, Anney R, Gallagher L, Heron EA. A genome-wide investigation into parent-of-origin effects in autism spectrum disorder identifies previously associated genes including SHANK3. *Eur J Hum Genet*. 2017;25(2):234–9 (**Epub 2016/11/24**).
- Mott R, Yuan W, Kaisaki P, Gan X, Cleak J, Edwards A, et al. The architecture of parent-of-origin effects in mice. *Cell*. 2014;156(1–2):332–42 (**Epub 2014/01/21**).
- Thorup AA, Jepsen JR, Ellersgaard DV, Burton BK, Christiani CJ, Hemager N, et al. The Danish High Risk and Resilience Study—VIA 7—a cohort study of 520 7-year-old children born of parents diagnosed with either schizophrenia, bipolar disorder or neither of these two mental disorders. *BMC Psychiatry*. 2015;15:233 (**Epub 2015/10/04**).
- Hemager N, Plessen KJ, Thorup A, Christiani C, Ellersgaard D, Spang KS, et al. Assessment of neurocognitive functions in 7-year-old children at familial high risk for schizophrenia or bipolar disorder: the danish high risk and resilience study VIA 7. *JAMA Psychiat*. 2018;75(8):844–52 (**Epub 2018/06/22**).
- Christiani CJ, Jepsen JRM, Thorup A, Hemager N, Ellersgaard D, Spang KS, et al. Social cognition, language, and social behavior in 7-year-old

- children at familial high-risk of developing schizophrenia or bipolar disorder: the Danish high risk and resilience study VIA 7—a population-based cohort study. *Schizophr Bull.* 2019;45(6):1218–30 (**Epub 2019/03/11**).
41. Burton BK, Thorup AAE, Jepsen JR, Poulsen G, Ellersgaard D, Spang KS, et al. Impairments of motor function among children with a familial risk of schizophrenia or bipolar disorder at 7 years old in Denmark: an observational cohort study. *Lancet Psychiatry.* 2017;4(5):400–8 (**Epub 2017/03/28**).
  42. Beauchaine TP, Constantino JN. Redefining the endophenotype concept to accommodate transdiagnostic vulnerabilities and etiological complexity. *Biomark Med.* 2017;11(9):769–80 (**Epub 2017/09/12**).
  43. Howey R, Cordell HJ. PREMIM and EMIM: tools for estimation of maternal, imprinting and interaction effects using multinomial modelling. *BMC Bioinformatics.* 2012;13:149 (**Epub 2012/06/29**).
  44. Henderson S, Sugden D, Barnett A. The movement assessment battery for children. 2nd ed. London: The Psychological Corporation; 2007.
  45. Cools W, Martelaeer KD, Samaey C, Andries C. Movement skill assessment of typically developing preschool children: a review of seven movement skill assessment tools. *J Sports Sci Med.* 2009;8(2):154–68 (**Epub 2009/01/01**).
  46. Wechsler D. Wechsler intelligence scale for children. 4th ed. San Antonio: The Psychological Corporation; 2003.
  47. White S, Hill E, Happe F, Frith U. Revisiting the strange stories: revealing mentalizing impairments in autism. *Child Dev.* 2009;80(4):1097–117 (**Epub 2009/07/28**).
  48. Constantino JN, Davis SA, Todd RD, Schindler MK, Gross MM, Brophy SL, et al. Validation of a brief quantitative measure of autistic traits: comparison of the social responsiveness scale with the autism diagnostic interview-revised. *J Autism Dev Disord.* 2003;33(4):427–33 (**Epub 2003/09/10**).
  49. Reynolds C, Kamphaus R. Reynolds intellectual assessment scales (RIAS). Lutz: Psychological Assessment Resources; 2003.
  50. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2014.
  51. Harrell FEJ, Dupont C. Hmisc: Harrell miscellaneous. 4.5-0 ed2021.
  52. Wei TS, Vilijam. R package 'corrplot': visualization of a correlation. 0.90 ed2021.
  53. Nudel R, Christiani CAJ, Ohland J, Uddin MJ, Hemager N, Ellersgaard D, et al. Quantitative genome-wide association analyses of receptive language in the Danish high risk and resilience study. *BMC Neurosci.* 2020;21(1):30 (**Epub 2020/07/09**).
  54. Nudel R, Christiani CAJ, Ohland J, Uddin MJ, Hemager N, Ellersgaard DV, et al. Language deficits in specific language impairment, attention deficit/hyperactivity disorder, and autism spectrum disorder: an analysis of polygenic risk. *Autism Res.* 2020;13(3):369–81 (**Epub 2019/10/03**).
  55. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559–75.
  56. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zonderman KT. Data quality control in genetic case-control association studies. *Nat Protoc.* 2010;5(9):1564–73 (**Epub 2010/11/19**).
  57. Abecasis GR, Cardon LR, Cookson WO. A general test of association for quantitative traits in nuclear families. *Am J Hum Genet.* 2000;66(1):279–92 (**Epub 2000/01/13**).
  58. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet.* 2002;30(1):97–101 (**Epub 2001/12/04**).
  59. van der Sluis S, Posthuma D. Single-locus association models. In: Ferreira MAR, Medland SE, Posthuma D, Neale BM, editors. *Statistical genetics: gene mapping through linkage and association*. London: Taylor & Francis; 2008.
  60. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Glied TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics.* 2010;26(18):2336–7 (**Epub 2010/07/17**).
  61. Ainsworth HF, Unwin J, Jamison DL, Cordell HJ. Investigation of maternal effects, maternal-fetal interactions and parent-of-origin effects (imprinting), using mothers and their offspring. *Genet Epidemiol.* 2011;35(1):19–45 (**Epub 2010/12/25**).
  62. Weinberg CR, Wilcox AJ, Lie RT. A log-linear approach to case-parent triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. *Am J Hum Genet.* 1998;62(4):969–78 (**Epub 1998/06/13**).
  63. Lange C, DeMeo DL, Laird NM. Power and design considerations for a general class of family-based association tests: quantitative traits. *Am J Hum Genet.* 2002;71(6):1330–41 (**Epub 2002/11/28**).
  64. Shim H, Chasman DI, Smith JD, Mora S, Ridker PM, Nickerson DA, et al. A multivariate genome-wide association analysis of 10 LDL sub-fractions, and their response to statin treatment, in 1868 Caucasians. *PLoS ONE.* 2015;10(4):e0120758 (**Epub 2015/04/22**).
  65. QTDT—online reference. <https://csg.sph.umich.edu/abecasis/QTDT/docs/index.html>. Accessed 30 March 2020.
  66. Engle RF. Chapter 13: wald, likelihood ratio, and Lagrange multiplier tests in econometrics. In: Leamer E, Heckman JJ, editors. *Handbook of econometrics*. Amsterdam: Elsevier; 1984. p. 775–826.
  67. Vösa U, Claringbould A, Westra H-J, Bonder MJ, Deelen P, Zeng B, et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *Nat Genet.* 2018. <https://doi.org/10.1101/447367>.
  68. GTEx Consortium. The genotype-tissue expression (GTEx) project. *Nat Genet.* 2013;45(6):580–5 (**Epub 2013/05/30**).
  69. Kamat MA, Blackshaw JA, Young R, Surendran P, Burgess S, Danesh J, et al. PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics.* 2019;35(22):4851–3 (**Epub 2019/06/25**).
  70. Stelzer G, Plaschkes I, Oz-Levi D, Alkelai A, Olender T, Zimmerman S, et al. VarElect: the phenotype-based variation prioritizer of the GeneCards Suite. *BMC Genomics.* 2016;17(Suppl 2):444 (**Epub 2016/07/01**).
  71. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: from gene data mining to disease genome sequence analyses. *Curr Protocols Bioinform.* 2016;54(1):1.30.1-1.3.
  72. Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martin D, et al. Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell.* 2016;167(5):1398–414 e24 (**Epub 2016/11/20**).
  73. Bonder MJ, Luijk R, Zhenakova DV, Moed M, Deelen P, Vermaat M, et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet.* 2017;49(1):131–8 (**Epub 2016/12/06**).
  74. Hubbard L, Tansey KE, Rai D, Jones P, Ripke S, Chambert KD, et al. Evidence of common genetic overlap between schizophrenia and cognition. *Schizophr Bull.* 2015;42(3):832–42.
  75. Sawa A, Snyder SH. Schizophrenia: diverse approaches to a complex disease. *Science.* 2002;296(5568):692–5 (**Epub 2002/04/27**).
  76. Kerner B. Toward a deeper understanding of the genetics of bipolar disorder. *Front Psych.* 2015;6:105 (**Epub 2015/08/19**).
  77. Stacey SN, Sulem P, Gudbjartsson DF, Jonasdottir A, Thorleifsson G, Gudjonsson SA, et al. Germline sequence variants in TGM3 and RGS22 confer risk of basal cell carcinoma. *Hum Mol Genet.* 2014;23(11):3045–53 (**Epub 2014/01/10**).
  78. Wu X, Cao W, Wang X, Zhang J, Lv Z, Qin X, et al. TGM3, a candidate tumor suppressor gene, contributes to human head and neck cancer. *Mol Cancer.* 2013;12(1):151 (**Epub 2013/12/03**).
  79. Bai Z, Stamova B, Xu H, Ander BP, Wang J, Jickling GC, et al. Distinctive RNA expression profiles in blood associated with Alzheimer disease after accounting for white matter hyperintensities. *Alzheimer Dis Assoc Disord.* 2014;28(3):226–33 (**Epub 2014/04/16**).
  80. Martin A, De Vivo G, Gentile V. Possible role of the transglutaminases in the pathogenesis of Alzheimer's disease and other neurodegenerative diseases. *Int J Alzheimer's Dis.* 2011;2011: 865432 (**Epub 2011/02/26**).
  81. Petrone A, Battaglia F, Wang C, Dusa A, Su J, Zagzag D, et al. Receptor protein tyrosine phosphatase  $\alpha$  is essential for hippocampal neuronal migration and long-term potentiation. *EMBO J.* 2003;22(16):4121–31.
  82. Heyes S, Pratt WS, Rees E, Dahimene S, Ferron L, Owen MJ, et al. Genetic disruption of voltage-gated calcium channels in psychiatric and neurological disorders. *Prog Neurobiol.* 2015;134:36–54 (**Epub 2015/09/20**).
  83. Dolphin AC.  $\beta$  subunits of voltage-gated calcium channels. *J Bioenerg Biomembr.* 2003;35(6):599–620.
  84. de Bagnaux PC, von Elsner L, Bierhals T, Campiglio M, Johannsen J, Obermair GJ, et al. A homozygous missense variant in CACNB4 encoding the auxiliary calcium channel beta4 subunit causes a severe neurodevelopmental disorder and impairs channel and non-channel functions. *PLoS Genet.* 2020;16(3):e1008625.

85. Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet.* 2011;43(10):969–76 (**Epub 2011/09/20**).
86. Havik B, Le Hellard S, Rietschel M, Lybaek H, Djurovic S, Mattheisen M, et al. The complement control-related genes CSMD1 and CSMD2 associate to schizophrenia. *Biol Psychiat.* 2011;70(1):35–42 (**Epub 2011/03/29**).
87. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet.* 2013;381(9875):1371–9.
88. Koiliari E, Roussos P, Pasparakis E, Lencz T, Malhotra A, Siever LJ, et al. The CSMD1 genome-wide associated schizophrenia risk variant rs10503253 affects general cognitive ability and executive function in healthy males. *Schizophr Res.* 2014;154(1–3):42–7 (**Epub 2014/03/19**).
89. Younis RM, Taylor RM, Beardsley PM, McClay JL. The ANKS1B gene and its associated phenotypes: focus on CNS drug response. *Pharmacogenomics.* 2019;20(9):669–84.
90. Carbonell AU, Cho CH, Tindi JO, Counts PA, Bates JC, Erdjument-Bromage H, et al. Haploinsufficiency in the ANKS1B gene encoding AIDA-1 leads to a neurodevelopmental syndrome. *Nat Commun.* 2019;10(1):3529.
91. Gros-Louis F, Dupre N, Dion P, Fox MA, Laurent S, Verreault S, et al. Mutations in SYNE1 lead to a newly discovered form of autosomal recessive cerebellar ataxia. *Nat Genet.* 2007;39(1):80–5 (**Epub 2006/12/13**).
92. Gama MTD, Braga-Neto P, Dutra LA, Alessi H, Maria LA, Gadelha AA, et al. Cognitive and psychiatric evaluation in SYNE1 ataxia. *Cerebellum.* 2019;18(4):731–7 (**Epub 2019/05/03**).
93. O’Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet.* 2011;43(6):585–9 (**Epub 2011/05/17**).
94. Piven J, Vieland VJ, Parlier M, Thompson A, O’Conner I, Woodbury-Smith M, et al. A molecular genetic study of autism and related phenotypes in extended pedigrees. *J Neurodevelop Disord.* 2013;5(1):30 (**Epub 2013/10/08**).
95. Yousaf A, Waltes R, Haslinger D, Klauck SM, Duketis E, Sachse M, et al. Quantitative genome-wide association study of six phenotypic subdomains identifies novel genome-wide significant variants in autism spectrum disorder. *Transl Psychiatry.* 2020;10(1):215 (**Epub 2020/07/07**).
96. St Pourcain B, Whitehouse AJ, Ang WQ, Warrington NM, Glessner JT, Wang K, et al. Common variation contributes to the genetic architecture of social communication traits. *Mol Autism.* 2013;4(1):34 (**Epub 2013/09/21**).
97. Li C, Dong S, Wang H, Hu Y. Microarray analysis of gene expression changes in the brains of NR2B-induced memory-enhanced mice. *Neuroscience.* 2011;197:121–31 (**Epub 2011/09/20**).
98. Loo SK, Rich EC, Ishii J, McGough J, McCracken J, Nelson S, et al. Cognitive functioning in affected sibling pairs with ADHD: familial clustering and dopamine genes. *J Child Psychol Psychiatry.* 2008;49(9):950–7 (**Epub 2008/07/31**).
99. Duvall JA, Lu A, Cantor RM, Todd RD, Constantino JN, Geschwind DH. A quantitative trait locus analysis of social responsiveness in multiplex autism families. *Am J Psychiatry.* 2007;164(4):656–62 (**Epub 2007/04/04**).
100. Coon H, Villalobos ME, Robison RJ, Camp NJ, Cannon DS, Allen-Brady K, et al. Genome-wide linkage using the social responsiveness scale in Utah autism pedigrees. *Mol Autism.* 2010;1(1):8 (**Epub 2010/08/04**).
101. Conklin HM. Verbal working memory impairment in schizophrenia patients and their first-degree relatives: evidence from the digit span task. *American J Psychiatry.* 2000;157(2):275–277. <https://doi.org/10.1176/appi.ajp.157.2.275>.
102. Haworth CM, Wright MJ, Luciano M, Martin NG, de Geus EJ, van Beijsterveldt CE, et al. The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Mol Psychiatry.* 2010;15(11):1112–20 (**Epub 2009/06/03**).
103. Jelenkovic A, Sund R, Hur YM, Yokoyama Y, Hjelmborg JV, Moller S, et al. Genetic and environmental influences on height from infancy to early adulthood: an individual-based pooled analysis of 45 twin cohorts. *Sci Rep.* 2016;6:28496 (**Epub 2016/06/24**).

## Publisher’s Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

