RESEARCH

Pleiotropy between language impairment and broader behavioral disorders—an investigation of both common and rare genetic variants

Ron Nudel[1,2,3] ![ORCID], Vivek Appadurai[1,2], Alfonso Buil[1,2], Merete Nordentoft[2,3,4] and Thomas Werge[1,2,4*]

Abstract

Background: Language plays a major role in human behavior. For this reason, neurodevelopmental and psychiatric disorders in which linguistic ability is impaired could have a big impact on the individual's social interaction and general wellbeing. Such disorders tend to have a strong genetic component, but most past studies examined mostly the linguistic overlaps across these disorders; investigations into their genetic overlaps are limited. The aim of this study was to assess the potential genetic overlap between language impairment and broader behavioral disorders employing methods capturing both common and rare genetic variants.

Methods: We employ polygenic risk scores (PRS) trained on specific language impairment (SLI) to evaluate genetic overlap across several disorders in a large case-cohort sample comprising ~13,000 autism spectrum disorder (ASD) cases, including cases of childhood autism and Asperger's syndrome, ~15,000 attention deficit/hyperactivity disorder (ADHD) cases, ~3000 schizophrenia cases, and ~21,000 population controls. We also examine rare variants in SLI/language-related genes in a subset of the sample that was exome-sequenced using the SKAT-O method.

Results: We find that there is little evidence for genetic overlap between SLI and ADHD, schizophrenia, and ASD, the latter being in line with results of linguistic analyses in past studies. However, we observe a small, significant genetic overlap between SLI and childhood autism specifically, which we do not observe for SLI and Asperger's syndrome. Moreover, we observe that childhood autism cases have significantly higher SLI-trained PRS compared to Asperger's syndrome cases; these results correspond well to the linguistic profiles of both disorders. Our rare variant analyses provide suggestive evidence of association for specific genes with ASD, childhood autism, and schizophrenia.

Conclusions: Our study provides, for the first time, to our knowledge, genetic evidence for ASD subtypes based on risk variants for language impairment.

Keywords: Specific language impairment, Autism spectrum disorder, Attention deficit/hyperactivity disorder, Schizophrenia, Polygenic risk score, Exome sequencing

Background

One of the most fundamental aspects of human behavior is communication through language. At the same time, it is also one of the most remarkable ones; children can acquire their mother tongue with ease and without conscious effort, and, yet, the mechanisms underlying this

*Correspondence: thomas.werge@regionh.dk
[4] Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
Full list of author information is available at the end of the article

Nudel *et al. Journal of Neurodevelopmental Disorders*   (2021) 13:54

Page 2 of 15

ability are largely unknown. There are many theories as to the nature and structure of human language, and they differ from one another from both the organizational and the representational perspectives. This, in turn, may have implications for accounts of language acquisition which use frameworks and concepts anchored in those theories [1]. However, from the molecular or genetic point of view, we do not need to presuppose much about the linguistic nature of the mechanisms which allow the child to acquire language or their relation to other cognitive domains. All the same, using genetics to investigate them may help answer some questions which pertain to higher levels of linguistic ability. Furthermore, it may also help answer questions which pertain to the links between linguistic ability and other behavioral or even physiological traits, in much the same way in which links between other behavioral and physiological traits and disorders have been found [2]. Thus, by exploring the genetic relationship between a primary form of language impairment and broader behavioral phenotypes, we could potentially identify pathways that may affect both language and other traits, which could, in turn, inform theories of language acquisition and development as well.

It has long been known, from twin studies and other family-based studies, that language ability and some language disorders are heritable [3]. For developmental spoken language disorders, pooling together twin data from across several studies obtained overall concordance rates of 83.6% for monozygotic twins and 50.2% for dizygotic twins, indicating a strong genetic component [3]. Twenty years ago, *FOXP2* became the first gene implicated in a speech and language disorder [4]. While the disorder was multi-faceted in terms of its behavioral phenotype, its genetic cause was a point mutation in a single gene. However, there are other disorders in which language is or may be impaired, and these may be complex, meaning that several genetic and environmental factors may combine to confer an increased risk of having them. Some complex disorders involve a child's broad behavioral neurodevelopment, and they include, among others, autism spectrum disorder (ASD) and attention deficit/hyperactivity disorder (ADHD), both of which are heritable [5] and may involve language deficits [6, 7]. In contrast, another disorder, namely, specific language impairment (SLI), is diagnosed when linguistic development is below age expectation in an otherwise typically developing child [8]. In recent years, the diagnostic criteria have changed, implementing a shift towards less exclusionary criteria and resulting in a new label: developmental language disorder[1], although ASD remained an exclusionary criterion [9]. Like ASD and ADHD, SLI is a complex disorder [10]. In contrast to the aforementioned monogenic speech and language disorder, *FOXP2* was not found to play a

major role in SLI susceptibility, thus suggesting a different genetic architecture and, perhaps, a more complex one [11]. Interestingly, several linguistic domains may also be impaired in schizophrenia [12], a psychiatric disorder not typically diagnosed in children, and language deficits in schizophrenia show familial aggregation [13]. Some studies implicated *FOXP2* either in either schizophrenia itself [14] or in language ability in schizophrenia patients [15], whereas other studies found no such associations [16, 17]. Of note, the SNP-based heritabilities of ASD, ADHD, and schizophrenia were recently estimated to be ~10%, ~20%, and ~13%, respectively, in the iPSYCH sample (which was also used in this study) [18].

Especially in the case of SLI and autism (autism being part of and arguably the core disorder within ASD), the (perhaps, superficial) similarities in linguistic impairment led to the question being raised of whether SLI and autism were on one continuum [19]. Until recently, most studies trying to answer this question focused on the linguistic deficits in SLI and autism, with some reporting similar linguistic domains being impaired in both disorders, and others reporting that children with SLI and children with autism are impaired on different domains [19–23], but it is generally said that the core deficits in SLI involve spoken language production and comprehension and the domains commonly affected in SLI are "structural" (phonology, morphology syntax and semantics), whereas the core linguistic deficit in autism affects mostly pragmatics (language use and hence social communication), although some children with SLI may exhibit some overlaps in the affected domains with children with autism, and vice versa [24]. What further complicated things was that some genes were linked to both disorders; for example, *CNTNAP2* (itself a *FOXP2* target [25]) was one of those genes [25, 26] (of note, it was also linked to schizophrenia [27]). Moreover, the top associations in the first genome-wide association study (GWAS) of SLI, in the model for child genetic effects, were with variants in genes previously implicated in ASD (and even in ADHD and schizophrenia) [28]; these included *CNTN5* [29], *RBFOX3* [30], and *THRB* [31, 32] (see Supplementary Table S1 for the top 10 SNPs from the discovery analysis from this study, i.e., with the updated dataset as per below, and corresponding genes). It was suggested that such observations (both linguistic and genetic) could be explained by a model incorporating genetic interactions, rather than only additive genetic overlaps; this type of model could account for genetic overlaps, while

---

[1] We use the term SLI in this paper, as the studies cited here are from before the change took place and used the more stringent definition and diagnostic criteria.

maintaining distinct linguistic profiles for SLI and autism [33]. It is worth mentioning that the reported genetic overlaps concerning specific genes were not always of the same nature across disorders; for example, common variants in *CNTNAP2* were associated with a language trait in children with SLI [25], but for ASD, a rare variant in that gene was also reported [34]. However, this is in line with theoretical accounts of disease-causing genes and also with experimental data showing associations with common variants in genes which are involved in related monogenic diseases [35]. This implies that both common and rare variants, the latter possibly having stronger deleterious effects, should ideally be examined when assessing genetic overlaps between disorders.

Polygenic risk scores as a tool for investigating cross-disorder genetic overlaps

A polygenic risk score (PRS) is an aggregate score that reflects an individual's genetic predisposition to a trait or a disease, as estimated based on prior genetic association data (typically from a GWAS for the given trait or disease). A PRS trained on a sample comprising cases and controls for a given disease is often used as a predictor for the same disease in an independent sample, but a PRS trained on one disease can be used to try to predict the risk of having another disease. This is known as a cross-disorder analysis and has been done for several psychiatric disorders; it provides a way of assessing genetic overlaps across disorders [36].

Polygenic risk scores in the clinical setting

Since their first use in a study of human disease [37], PRSs have become a popular tool in research. The nature of the PRS, i.e., being one aggregate score capturing an individual's genetic predisposition to having a particular trait, also means it can be readily used by researchers in disciplines typically far removed from genetics, such as the social sciences, where it can be incorporated into statistical models, thus allowing the integration of genetic information and social outcomes [38]. But although PRSs have been successfully used in research contexts, typically allowing to differentiate cases and controls at the group level, they cannot, as of now, be used as predictors for disease at an individual level [39]. Nonetheless, integrating PRSs into the clinical setting remains one of the main goals of PRS research, and, even though a PRS cannot stratify individuals from the entire population based on the individual probabilities of their developing a disease, it could, together with clinical risk factors, potentially help identify a group of individuals with a particularly high risk for some diseases [40]. One example of a successful application of this approach was using a PRS to identify individuals at high risk of coronary artery disease; using a PRS, the authors were able to identify a group of individuals (8% of the population) with high risk of developing coronary artery disease, with an odds ratio ≥ 3 [41].

Polygenic risk scores in studies of language-related traits

In psychiatry, where the clinical presentation of various conditions might be more complex, PRSs could be used for distinguishing subtypes of psychiatric disorders. A PRS for schizophrenia has been shown to differentiate schizoaffective bipolar disorder cases from the rest of the bipolar disorder sample in one study [42]. However, when it comes to prognostic value, a recent study found no significant improvement in using a PRS for schizophrenia when predicting poor outcomes (proxies for a poor clinical trajectory, including: aggressive behavior, requiring in-patient psychiatric treatment, prescription of two or more unique antipsychotics, prescription of clozapine, self-harm and homelessness), relative to current standards of care [43], even for models in which the PRS was significantly associated with the proxy (i.e., it significantly explained some of the variance in those traits), which was the case for the first two out of the above six outcomes. This example illustrates the fact that, even if the PRS is significantly associated with an outcome, it does not mean that adding it to the prediction model would improve the performance of the model relative to including only clinical features.

As in the general case of psychiatric disorders described above, studies of language-related disorders are also plagued by the heterogeneity of the disorders, which, in turn, may also influence clinical diagnosis and treatment, even in terms of access to support services in the first place [44]. In this respect, studying the genetic architecture of the disorder could be informative as to the boundaries (and similarities) between disorders with overlapping phenotypes. Studies applying PRSs to language-related traits or disorders in the clinical context are scarce, but attempts have been made to investigate the potential use of PRSs in these settings. A PRS based on several language measures was shown to explain a small proportion of the variance in language and psychosocial problems in 8-year-old children, although this PRS was not genome-wide and consisted of markers in preselected candidate genes [45]. Another recent study examined the potential application of a genome-wide PRS for educational attainment to identifying children with language and literacy problems at an early developmental stage. The PRS for educational attainment significantly explained a small proportion of the variance of language and literacy at age 12, but its predictive ability was overall low and deemed not useful in the clinical setting [46]. In the neighboring field of speech and voice

disorders, the use of genetic information in the clinical setting has also been advocated [47]. In addition to investigating the direct association between PRS for a relevant trait and developmental outcomes, clinical practice may also benefit, albeit not immediately, from studies into the genetic overlaps across language-related disorders and traits. For example, previous investigations which used PRSs in a cross-disorder setting identified genetic overlaps between ADHD and reading-related traits (showing a negative association) [48]. Interestingly, these traits also showed a positive association with PRSs for educational attainment in the same study.

In our own previous study (hereafter referred to as the pilot study), which included a family-based cohort comprising children assessed for language, intelligence, and other behavioral traits through test batteries and interviews, we used a PRS trained on the SLI GWAS in trying to predict risk of ASD and ADHD, using case-control datasets for these disorders from among the unrelated children of the cohort ($N = 391$). In our study, we observed that, overall, the PRS significantly predicted some risk of SLI, used as a positive control in the target sample, but it did not predict risk of ASD (mutually exclusive with SLI) or ADHD—or height, used as a negative control [49]. Thus, at least when it came to common variants, we did not observe a genetic overlap between SLI and ASD, or between SLI and ADHD, as captured by a genome-wide PRS. The biggest limitations of the previous study were the sample size and the fact that the ASD diagnosis encompassed children with varying language profiles (as determined from their performance on a receptive language test). The aim of this extended study is thus fourfold: (i) to apply SLI-trained PRS to a much larger sample, which comprises more than ten thousand cases each of ASD and ADHD; (ii) to see whether different results are obtained when examining childhood autism and Asperger's syndrome (which differ in their linguistic profiles, in this case, based on International Classification of Diseases (ICD) criteria) separately, and to assess whether this could potentially be used to guide clinical diagnosis; (iii) to extend the analysis to include schizophrenia; (iv) to examine potential genetic overlaps between SLI and the other disorders using exome sequencing data which include rare variants in SLI candidate genes and other language-related genes.

Methods

Study population and phenotypes

The individuals in this study are part of the Danish iPSYCH case-cohort sample [50], which comprises individuals selected either for having at least one of six disorders (ASD, ADHD, schizophrenia, bipolar disorder, depression and anorexia) or as part of a random population sample. The iPSYCH samples underwent extensive quality control (QC) procedures based on both genetic data and registry data to remove ancestry outliers, duplicate samples, individuals with cryptic relatedness, and individuals with low-quality genotype measures, as described in an earlier study [18]. This resulted in a sample of 65,534 unrelated Danish individuals, as used in previous studies [51–53]. The phenotypes used in this study include the 2016 dataset of diagnoses from the Danish Psychiatric Central Research Register for these 65,534 individuals. The diagnoses correspond to the following ICD-10 [54] codes: ASD (F84.0, F84.1, F84.5, F84.8, and F84.9), childhood autism (F84.0), Asperger's syndrome (F84.5), ADHD (F90.0), and schizophrenia (F20). Equivalent ICD-8 [55] codes might have been used for schizophrenia (295.x9 excluding 295.79) and childhood autism (299.00), depending on when the individual received the diagnosis. For each phenotype, cases were defined as having the respective diagnosis as per the above codes, and controls were defined as (i) not having the diagnosis in question and (ii) having been included in iPSYCH as part of the random population sample, i.e., an individual who is included only in the case subset of iPSYCH will not be included as a control for another case diagnosis (which they do not have), but they are considered a case for the diagnosis they do have. Individuals in iPSYCH may be cases for more than one disorder.

Genetic quality control

The samples were genotyped on the Illumina PsychArray v1.0. Preliminary QC steps on the raw genotype data (based on call rates and the Gentrain score) are described in the original iPSYCH paper [50] and subsequent QC is described in a later iPSYCH study [56]. The marker dataset used in this study was filtered further with PLINK [57] v1.90b3o to remove markers with rare variants and non-autosomal markers, and later with v1.90b3.34 to remove one marker from every pair of markers with duplicate positions. The final dataset had markers with a minimum minor allele frequency (MAF) of 0.009632 and maximum missingness of 0.01257. All but 218 markers had Hardy-Weinberg equilibrium p value $> 1 \times 10^{-6}$ in controls. We report these numbers and not thresholds, as the QC steps were performed in a larger subset of the iPSYCH sample (a homogeneous sample of European ancestry) than that used in this study, which was selected for the purpose of QCing the markers prior to imputation for another study. Further details are given in that study [18] and in a subsequent study [58]. In total, 242,077 markers were retained following these steps, and 239,582 markers remained after removing markers from the major histocompatibility complex (MHC) region.

Nudel *et al. Journal of Neurodevelopmental Disorders*   (2021) 13:54

Page 5 of 15

Note that the above QC describes the marker QC; the final sample used in this study comprised only individuals passing the sample QC as mentioned in the previous section and as described in [18].

Polygenic risk scores and regression models

The summary statistics used in the construction of the PRS were taken from an updated analysis of a previous SLI GWAS [28]. To our knowledge, this is currently the only GWAS of SLI, and it is based on the largest available sample of SLI families, the SLI Consortium sample. The dataset used in this study was the most strictly QCed one (termed "Correction 1"), as used in our pilot study, a recent study of PRS in SLI, ASD, and ADHD, which details the complete protocol for this dataset [49]. In short, the SLI phenotype in the discovery study was based on proband status and/or low receptive or expressive language scores from a standardized test. The SLI Consortium also employed exclusion criteria which included low non-verbal intelligence and/or an indication of autism, as detailed in the original papers [28, 59–61]. The average numbers of family subsets per single-nucleotide polymorphism (SNP) in the updated GWAS were as follows: 150 case-parents trios, 55 case-mother duos, 12 case-father duos, and 19 cases (and sometimes case parents, but these were few, on average < 1 per SNP). These subsets were generated per marker by the PREMIM tool, which generates the input to EMIM (the software with which the GWAS was performed) [62], in a way that prioritizes case-parents trios. For example, if for a given marker and for a given family both parents and a case have genotypes, then this would be a trio subset. If the paternal genotype is missing for this marker, then this would be a case-mother duo, and so on. The GWAS was family-based (not case-control; only case subsets were used as per the above), whereby the effect estimated for each SNP in the model used in the GWAS was one effect or risk parameter, R_1 (defined as the factor by which a child's disease risk is multiplied if they possess one risk allele), so that the increase in risk from carrying two risk alleles was defined as the square of R_1 [63]. By default, the minor allele was defined as the "risk" allele, but it could also be protective, in which case the effect parameter would be < 1. These effects were used in the calculation of the PRS, similar to the use of odds ratios (ORs). SNPs which had a "warning" value of 1 from EMIM (i.e., there was some problem in the models for those SNPs) were removed from the summary statistics. Further information about the model employed in the discovery GWAS can be found in the supplementary notes for this paper. PRSs were generated for iPSYCH individuals using PRSice v2.2.6 [64] with the following clumping parameters: r^2 value of 0.2 in a 500-kbp window, as

recommended for psychiatric traits [36], and a *p* value threshold of 1, both to conform to the protocol used in the pilot study and to increase the accuracy of the PRS (as it has been observed that, when the discovery sample is not very large, including all SNPs can lead to better performance, and both experimental and simulation studies reported better performance when including all SNPs in most cases; this is particularly applicable to cases in which the original GWAS did not identify many genome-wide significant associations) [36, 49, 65, 66]. As before, SNPs from the MHC region and ambiguous (A/T and G/C) SNPs were excluded. For binary traits, the program performs a logistic regression of the phenotype on the PRS and outputs Nagelkerke's R^2 as well as an adjusted R^2 (the adjustment is for the proportion of cases and prevalence for each phenotype) [67]; to that end, the following prevalence values were provided to PRSice for the ASD, ADHD, and schizophrenia phenotypes, respectively: 1% [68], 5% [69], and 0.4% [70]. In the analyses for the ASD subtypes, the prevalence value used for childhood autism was 0.4% [68], and for Asperger's syndrome, it was 0.3% [71]. Otherwise, the default parameters of PRSice were used. The logistic regressions of the phenotype on the PRS were also repeated in R v3.3.1 [72] using PRS scaled across the entire sample (using the *scale* function in R with the default parameters), so that the regression odds ratios are derived from coefficients corresponding to a change of 1 standard deviation (SD) in the PRS, as presented in the "Results" section. The reported two-sided *p* values for the models are for these coefficients' being different from zero, as evaluated by the function in R (using the *t*-distribution for a linear regression (using the *t*-statistic), e.g., for height in the pilot study, and the normal distribution for a logistic regression (using the Wald *z*-score), e.g., for ASD, as implemented in the *lm* and *glm* functions in R, respectively). Confidence intervals (CIs) for the coefficients were estimated using the *confint* function. The sample sizes (cases; controls) for the PRS analyses were as follows: ASD (12,884; 21,321), childhood autism (3,313; 21,634), Asperger's syndrome (4,710; 21,567), ADHD (15,060; 21,265), schizophrenia (2,867; 21,596).

Candidate genes for language disorders and traits and rare variant group tests

Since only a very small proportion of the SLI Consortium proband sample was exome-sequenced [73], and given the reported genetic overlaps between monogenic disorders and common variants in phenotypically related disorders, as discussed earlier, we included candidate genes implicated through both common and rare variants in the exome-sequencing analyses. As a first step, we used recent review articles about the genetics

Nudel *et al. Journal of Neurodevelopmental Disorders*   (2021) 13:54

Page 6 of 15

of language disorders and related conditions [74, 75], as well as a survey of some of the literature from our recent work on receptive language [76], to identify studies in which at least one of the investigated phenotypes was spoken language impairment or a spoken language trait. We then included two categories of genes in the rare variant analyses: (i) genes implicated directly in spoken language impairment: *CNTNAP2* [25], *CMIP*, and *ATP2C2* [77], *NOP9* [28], *NFXL1* [78], *SETBP1* [79], *NDST4* [80], and *OXR1*, *MUC6*, *SCN9A*, *FAT3*, *KMT2D*, and *PALB2* [73]; (ii) genes implicated in studies of spoken language traits in a general population sample not selected for having low language ability: *ABCC13* [81] and *RORB* [82]. Human leukocyte antigen (HLA) genes were not included due to the complex genetic architecture of the MHC region and the fact that they (and their overlaps across disorders) had already been extensively examined in past studies of SLI, ASD, ADHD, and schizophrenia [52, 83–87]. While additional genes have been implicated in broader disorders or phenotypes involving language, we chose to keep genes reported specifically for spoken language impairment or spoken language traits not in combination with other traits (e.g., not language impairment and reading impairment modeled simultaneously or speech-related disorders, and so on). Additionally, we required that the gene be implicated directly, that is, through a gene-based analysis, or, in case of an association study, that the associated markers be within the gene. This was done to ensure that the rare variant analyses are closer to the PRS analyses (which were based on common variants for SLI, i.e., spoken language impairment)—even though the two approaches differ in methodology and interpretation—and in order to be able to draw conclusions regarding the potential overlaps between spoken language disorders/traits proper and the other phenotypes. Genes were selected from the above studies based on reported significance levels within each study, or on the gene being the top candidate in a given study based on its *p* value or qualitative measures, e.g., genes with co-segregating rare variants or genes highlighted through compound heterozygous inheritance in the exome-sequencing study of SLI. A flowchart summarizing the selection process can be found in Supplementary Figure S1. The starting point for the rare variant analyses was a dataset generated for, and described in detail in, a recent iPSYCH study [88]. The genetic data (exome variants) for this dataset were generated independently of the genotype array data described earlier in the "Methods"; however, individuals failing the iPSYCH sample QC, e.g., on account of having non-Danish ancestry, were excluded from the pedigree/phenotype file provided to the program which performed the tests, so that every individual in the exome-sequencing dataset must also have passed the general iPSYCH sample QC as referenced above. The genomic coordinates for the above list of genes were obtained from Release 19 (GRCh37.p13) of GENCODE, and variants in those positions were extracted from the iPSYCH VCF file using BCFtools v1.9 [89]. The new VCF file was annotated using snpEff v4.3t [90] with the GRCh37.p13 database. The variants kept for downstream analysis were of the following types: frameshift variants; missense variants; nonsense (stop gained) variants; splice site donor, acceptor, or region variants, all with a maximum MAF of 1%. The statistical test employed was the optimized SKAT test (SKAT-O) [91] as implemented in EPACTS v3.2.6 (with the default parameters, apart from the maximum MAF as per the above) [92], for which the variants in each gene were grouped together. SKAT-O optimally combines the burden test (which collapses the variant counts for all markers in a region) and the SKAT (which sums up the squares of the variant score statistics for all markers in a region), both of which examine aggregate variant effects, but perform optimally in different scenarios; the burden test is most suitable when most variants in a region are causal and their effects are in the same direction, and SKAT is most suitable when a large proportion of the variants in a region are either non-causal, or have effects that are in different directions [91]. The *p* value for the test is for the enrichment of rare variant associations per gene, and the ratio Rho reflects the optimal combination of the two kinds of tests (1 corresponds to a pure burden test, and 0 corresponds to a pure SKAT). No variants (passing QC) were found in *CNTNAP2*, and the gene was therefore not included in the tests. The sample sizes (cases; controls) for the rare variant analyses were as follows: ASD (9,579; 8,782), childhood autism (2,343; 8,987), Asperger's syndrome (3,482; 8,944), ADHD (7,396; 8,816), schizophrenia (1,980; 8,968).

Comparison with the pilot study

We present results from analyses which used the pilot study [49] sample (run with PRSice v2.2.3) for comparison with the current study, as these results and their contrast or similarity with the results from the iPSYCH sample are important for the interpretation of the findings from the present study. The pilot study sample consisted of unrelated children who were part of a family-based study, the Danish High Risk and Resilience Study – VIA 7, who were assessed for language performance, intelligence, and other behavioral traits [93], as detailed in our previous publications which used the genetic data for these children [49, 76]. The pilot study paper details the sample size and the criteria for the affection status or measurement for each phenotype included here (note that SLI was also termed "narrow language phenotype"

Nudel *et al. Journal of Neurodevelopmental Disorders*   (2021) 13:54

Page 7 of 15

Table 1 Results of the PRS regression analyses in iPSYCH; updated results from the pilot study sample are shown for comparison

Phenotype	R^2/Nagelkerke's R^2 (%)	Adjusted R^2 (%)	Scaled PRS beta/ odds ratio (95% CI)	P value
iPSYCH sample*				
ASD	*0.017*	*0.007*	*1.02 (1.00–1.05)*	*0.037*
Childhood autism	**0.079**	**0.042**	**1.06 (1.02–1.10)**	**0.001**
Asperger's syndrome	0.002	0.001	1.01 (0.98–1.04)	0.552
ADHD	0.002	0.002	1.01 (0.99–1.03)	0.426
Schizophrenia	0.002	0.001	0.99 (0.95–1.03)	0.613
Pilot study sample**				
ASD	0.017	0.017	0.97 (0.59–1.57)	0.887
ADHD	0.008	0.009	1.02 (0.75–1.39)	0.897
SLI	3.5	5.09	1.60 (1.02–2.54)	0.041
Height	0.006	NA	0.07 ((− 0.98) –1.13)	0.895

The adjusted R^2 includes an adjustment for the prevalence and proportion of cases in each sample, as implemented in the PRSice software. Associations in boldface survive Bonferroni correction for multiple testing (not applicable to positive or negative controls). Associations in italic are only nominally significant (not applicable to positive or negative controls)

PRS polygenic risk score, *CI* confidence interval, *ASD* autism spectrum disorder, *ADHD* attention deficit/hyperactivity disorder, *SLI* specific language impairment, *NA* not applicable

The R^2 reported for height is a true R^2 from a linear regression and not a Nagelkerke's R^2, and the effect is the beta (regression coefficient) and not an odds ratio; it should therefore be judged relative to 0 and not to 1. The height variable was collected in different ways at the Copenhagen and Aarhus centers for the VIA 7 study; the Aarhus center used, for the most part, a uniform method of measuring height with a special device, but we note that the PRS is not significantly predictive of height even if we restrict the analysis to include only children measured in Aarhus. We include this for the sake of completeness, even though the sample size is much smaller and thus the estimates are less reliable ($R^2 = 0.7\%$, $P = 0.419$)

*72,932 SNPs were included in the PRS

**80,121 SNPs were included in the PRS

in that paper). We include updated and slightly different results here as compared to the pilot study as published, as subsequent QC in the family-based sample revealed some Mendelian errors in child-parent duos not previously identified (as duos are not checked by default by PLINK), as reported in a subsequent study which used the same sample [76]. This did not result in the removal of any duos that were not removed at a later stage anyway (in the relatedness check), but a number of markers and genotypes were removed (the conclusions of the original study were not affected by this). For the binary traits in the pilot study, the regression models were the same as those described for iPSYCH, with the same prevalences as mentioned earlier for ASD and ADHD, and a prevalence of 7% for SLI [94]. For height, a linear regression was performed with covariates for the age at measurement and sex. The reported R^2 for the PRS was calculated as the R^2 for the full model (height regressed on PRS and covariates) minus the R^2 for the null model (height regressed on the covariates), as implemented in PRSice. Lastly, we report some new analyses which used the pilot study sample but were not included in the pilot study, as they are relevant for comparison with the iPSYCH sample.

Difference in PRS between childhood autism cases and Asperger's syndrome cases

Following the regression analyses and in order to evaluate the difference in PRS between childhood autism and Asperger's syndrome, we performed a Mann-Whitney U test with the *wilcox.test* function in R, using the scaled PRS. We performed a one-sided test, as we expected cases of childhood autism to have a higher PRS than cases of Asperger's syndrome. For this purpose, we excluded a small number of children who had both diagnoses ($N = 175$). Area under the curve (AUC) values were computed with the *auc* function of the pROC package v1.17.0.1 in R [95]. For the purpose of calculating the AUC, childhood autism cases were defined as "cases" (affection status 1) and Asperger's syndrome cases were defined as "controls" (affection status 0).

Results

The results of all PRS analyses are shown in Table 1. Overall, the SLI-trained PRS, which was previously found to be predictive of SLI in an independent sample in our pilot study, was not predictive of the risk of ASD or ADHD (adjusted R^2 close to 0%, neither of them significant after Bonferroni correction, $N = 5$), in line with and thus replicating the results of our pilot study; for the additional phenotype of schizophrenia, the PRS was not predictive. While the result

Table 2 Results of group comparisons using the Mann-Whitney *U* test or AUC analyses

Group 1	Group 2	Type of analysis	Result	*P* value
Childhood autism cases (iPSYCH sample)	**Asperger's syndrome cases (iPSYCH sample)**	**Mann-Whitney *U* test**	**$W = 7,353,100$; difference in location = 0.059**	**0.006**
Childhood autism cases (iPSYCH sample)	Asperger's syndrome cases (iPSYCH sample)	AUC	0.5167	NA
SLI cases (pilot study sample)	SLI controls (pilot study sample)	AUC	0.6276	NA
SLI cases (pilot study sample)	ASD cases (pilot study sample)	AUC	0.6303	NA

Associations in bold survive Bonferroni correction for multiple testing

AUC area under the curve, *NA* not applicable, *ASD* autism spectrum disorder, *SLI* specific language impairment

for ADHD was similar in terms of effect size, R^2 and *p* value in both studies, the result for ASD was not. The explanatory power of the PRS was close to zero in both cases (~0.02% in the pilot study; ~0.01% in the current study), but the association in the pilot study was in the opposite direction compared to the new result for ASD, and the latter was at least nominally significant, unlike the former result. Since the current sample is much larger, the effect estimate is more accurate, and the confidence interval is smaller. This new result suggests that, while, by large, the genetic overlap (from common variants) between SLI and ASD is small, it may nonetheless be different from zero, and that at least some of the overlapping loci have effects in the same direction. This is illustrated more strongly when comparing the models for childhood autism and Asperger's syndrome: even though there were fewer cases of childhood autism compared to both ASD in general and Asperger's syndrome in particular, the model for childhood autism performed better, with R^2 ≈ 0.04% and $P = 0.001$, which survives Bonferroni correction for multiple testing ($N = 5$), whereas the model for Asperger's syndrome was not predictive. The above models tested each ASD subtype against controls; we therefore sought to evaluate the difference between the two case groups directly. Our Mann-Whitney *U* test found a significant difference between the childhood autism and Asperger's syndrome case groups ($W = 7,353,100$, difference in location = 0.059, $P = 0.006$, lower bound of a 95% confidence interval (CI) = 0.02). This corresponds to an AUC of ~52%, which is only ~2% over what is considered completely uninformative. Of note, using the same approach with the updated dataset from the pilot study, we obtain an AUC of ~63% for SLI cases *versus* SLI controls (one-sided $P = 0.025$ for the *U* test), and a similar AUC of ~63% for SLI cases *versus* ASD cases ($P = 0.09$). These results are summarized in Table 2.

The results of the rare variant analyses are shown in Table 3. Tests for three genes obtained nominally significant *p* values: *NDST4* in ASD, *RORB* in childhood autism, and *SETBP1* in schizophrenia. However, none of these survive Bonferroni correction for multiple testing ($N = 70$).

Discussion

Our extended study replicated the results of our pilot study, namely, that, overall, there does not seem to be statistically significant genetic overlap between SLI and ASD or ADHD. However, the degree of overlap between SLI and ASD was determined more accurately in this study, and the new result was nominally significant, before correction for multiple testing ($P = 0.037$). Moreover, we observed a difference between the model for childhood autism and the model for Asperger's syndrome in terms of the predictive ability of the PRS, suggesting some pleiotropy between childhood autism and SLI: while still explaining only a small proportion of the risk of childhood autism, the SLI-trained PRS nonetheless achieved statistical significance (surviving Bonferroni correction) only in the former case. The analyses in Table 1 for these two phenotypes test each case group against controls, which may be shared between the two case-control datasets and hence are not independent. When we test both case groups against each other, we find a significant difference, with a tendency for the childhood autism group to have a higher SLI-trained PRS. Keeping in mind that the PRS represented log-additive genetic risk of language impairment, this result shows an intriguing correspondence between this genetic difference between childhood autism and Asperger's syndrome, which could be seen as a difference in the "genetic load for language impairment," and the language profiles of the two disorders, which constitute the major difference between them [96]: ASD includes a group of pervasive developmental disorders involving abnormal social interaction, abnormal behavior patterns (typically involving restricted, stereotyped and repetitive behavior), and impaired communication [6]. Childhood autism is characterized by deficits in all of the above domains, while a diagnosis of Asperger's syndrome is typically given when there are no evident communication deficits or language delay. In this study, the

Nudel *et al. Journal of Neurodevelopmental Disorders*   (2021) 13:54

Page 9 of 15

Table 3 Results of the rare variant analyses

Phenotype	ASD			Childhood autism			Asperger's syndrome			ADHD			Schizophrenia		
Gene	Number of variants	P value	Rho	Number of variants	P value	Rho	Number of variants	P value	Rho	Number of variants	P value	Rho	Number of variants	P value	Rho
NOP9	61	0.059	0	46	0.103	0	49	0.108	0.5	57	0.242	0	47	0.111	0.4
KMT2D	454	0.227	0	361	0.545	0	361	0.319	1	435	0.533	1	337	0.126	0.7
RORB	14	0.243	1	11	0.033	1	12	0.894	1	14	0.300	1	11	0.397	1
ABCC13	1	0.138	NA	1	0.180	NA	1	0.584	NA	1	0.383	NA	1	0.079	NA
NDST4	45	0.011	0.4	42	0.270	1	43	0.117	0.4	48	0.215	1	44	0.142	1
FAT3	361	0.361	1	289	0.549	1	291	0.420	1	337	0.144	0.6	278	0.726	0
CMIP	22	0.402	1	18	1.000	1	18	0.637	0	23	1.000	0	17	0.715	0
PALB2	83	0.551	0	64	0.592	0	67	0.827	1	75	0.747	1	65	0.633	0
SCN9A	130	0.093	0	103	0.578	0	109	0.396	0	126	0.153	0	102	0.308	1
MUC6	31	0.857	1	27	0.816	0	28	1.000	1	32	0.752	1	28	0.815	1
NFXL1	26	0.519	1	13	0.695	0	15	0.235	1	17	0.849	1	13	0.559	0
OXR1	41	0.572	0.7	34	0.876	1	34	0.779	1	41	0.365	0	34	0.497	1
ATP2C2	161	0.895	1	122	0.460	0	131	0.317	1	159	0.359	1	118	0.650	1
SETBP1	106	0.777	1	88	0.790	1	91	0.365	1	105	0.322	1	85	*0.018*	*1*
CNTNAP2	0	NA	NA	0	NA	NA	0	NA	NA	0	NA	NA	0	NA	NA

The number of variants refers to the number of variants passing QC and count/frequency thresholds for each gene. Associations in italic are only nominally significant

ASD autism spectrum disorder, *ADHD* attention deficit/hyperactivity disorder, *NA* not applicable

childhood autism and Asperger's syndrome diagnoses followed the ICD guidelines (almost exclusively ICD-10 for the former, and only ICD-10 for the latter). The ICD-10 criteria for childhood autism specify, among other things: the characteristic type of abnormal functioning in all the three areas of psychopathology: reciprocal social interaction, communication, and restricted, stereotyped, repetitive behavior. For Asperger's syndrome, it states: it differs from autism primarily in the fact that there is no general delay or retardation in language (https://icd.who.int/browse10/2019/en#/F84, accessed May 16, 2021). It should nonetheless be acknowledged that there may yet be some language problems associated with Asperger's syndrome, too [97, 98], only not to the same extent as in childhood autism, and some theorize that childhood autism and Asperger's syndrome are quantitatively different, rather than qualitatively different [96]. We also observed that a small number of children seemed to "transition" from one diagnosis to the other, or get both codes, although this could be the result of a misdiagnosis or some other kind of error inherent to registry-based research. In summary, the results of our PRS analyses indicate a subtle, but statistically significant, difference in the genetic load for language impairment between childhood autism and Asperger's syndrome. This suggests that, at least at the group level, these two ASD subtypes can be distinguished by their genetic risk of language impairment, although the difference is very small, as reflected in the AUC.

In the rare variant analyses, three genes were nominally significantly associated with a disorder. Variants in *NDST4* were associated with ASD, variants in *RORB* were associated with childhood autism, and variants in *SETBP1* were associated with schizophrenia. *NDST4* was included due to its implication in language impairment [80]. The gene belongs to a family of genes called GlcNAcN-deacetylase/N-sulfotransferases, which have important roles in development [99]. While its connection to language is not clear, it has been associated with traits such as drinking behavior [100] and circulating levels of resistin, a hormone involved in inflammation [101], and some protein-truncating variants have been reported for this gene in the context of schizophrenia [102]. *RORB* was included due to its association with verbal intelligence, namely, with a vocabulary measure [82]. The protein encoded by this gene is a nuclear receptor [103] and it has been implicated in bipolar disorder [104]. Given its association with vocabulary, it is not surprising that it should show some association with childhood autism, as one study showed poor vocabulary growth to be associated with autism severity at 6 months from the start of the study (the participants' initial chronological ages were 20–71 months) [105]. Notably, this gene has

been highlighted in a recent ASD exome-sequencing study which included the iPSYCH sample but used a different methodology [106]. Lastly, *SETBP1* was included due to its implication in language impairment [79]. This gene is a transcription regulator [107], and it has been implicated in several studies of related disorders, such as childhood apraxia of speech [108] and developmental delay/expressive language delay [109–111]. There is some new evidence for its involvement in schizophrenia in a recent study [112], and it was also significant in some of the analyses of an exome-sequencing study of ASD [106].

Limitations of our study

The training dataset for the PRS used in this study is, to the best of our knowledge, the only GWAS of SLI to date. As the primary sample was collected about 20 years ago and was originally intended for linkage analyses, it consists mainly of families of SLI probands, where unaffected individuals are related to affected individuals. The GWAS sample included several hundred individuals in subsets of case-parents trios, case-parent duos, cases, and so on, and, in the specific family-based GWAS design employed, only case subsets were used (i.e., controls were not used in the association tests themselves). As such, this analysis is inherently different from a standard case-control GWAS. While the SLI Consortium sample is not large by today's standards of case-control studies, it is not atypical for family-based genetic studies. Another limitation is that the iPSYCH sample had no SLI phenotype or any kind of standardized language test score. However, the results from the pilot study sample that we obtained for our positive control (SLI) both in terms of the Nagelkerke's R^2 and the adjusted R^2 were, in fact, higher (Table 1)[2] than the maximum value obtained for schizophrenia (3.2%) in the study which conceptualized PRS analyses for human disease [37]. For schizophrenia, the R^2 rose to 18.4% with a much larger discovery dataset (from a meta-analysis) a few years later [113], but a similar meta-analysis is currently not feasible for SLI. In summary, one limitation of our study is the sample size of the original GWAS, although it should be emphasized that we employed several tests and controls to assess whether the PRS predicts what it is supposed to predict, and we followed the conventional guidelines for PRS analyses in which the discovery sample was small, as explained in the "Methods" section. It should also be mentioned that, even though the R^2 in the aforementioned original schizophrenia study was lower than in our study, the PRS

---

[2] The values rise to Nagelkerke's $R^2 = 3.87\%$ and adjusted $R^2 = 6.05\%$ ($P = 0.024$, with the updated dataset), when less stringent criteria are applied to SLI controls (i.e., when controls have language test scores above the threshold defining cases (1.5 SD below the population mean), as opposed to above a score defined as 0.5 SD below the population mean, as used in the main analyses), similar to what has been shown previously [49]).

Nudel *et al. Journal of Neurodevelopmental Disorders*   (2021) 13:54

Page 11 of 15

it was based on was nonetheless used in a cross-disorder analysis, much like in this study.

A limitation in terms of the applicability of the results is that the difference in the SLI-trained PRS between childhood autism cases and Asperger's syndrome cases was not large enough for clinical utility; the AUC for this model was too small for this at this stage, but, as proof of concept, our results are nonetheless promising; they suggest three things: (i) that, as observed in the pilot study, an SLI-based PRS is not a good predictor of ASD, meaning that the genetic correlation between the disorders is not expected to be large; (ii) that there is, however, a small but significant positive genetic overlap between SLI and childhood autism in particular, meaning that some loci could be shared between the two disorders (which could potentially be many loci with small effects)—these two results can inform us on the relationship between SLI and ASD and childhood autism; and (iii) that those overlapping loci could potentially distinguish between two types of autism spectrum disorder, one in which language is typically impaired, and another in which it is not. Given adequate training sets and sample sizes, this could, in the future, lead to a way of distinguishing between subtypes of ASD using genetic risk scores trained on language impairment. In the rare variant analyses, none of the tests survived a Bonferroni correction for multiple testing, and, therefore, they can provide at most suggestive evidence for association at this stage. This could be due to lack of power, as only a subset of the iPSYCH sample was exome-sequenced, and, by definition, rare variants are found in low numbers across samples. It should be noted that, when genetic correlation is not observed (i.e., even if it equals or is close to zero), it does not mean that pleiotropy does not exist, as the former depends on the directionality of the effects [2, 114]. In this context, it is worth noting that PRS cross-disorder analysis typically agrees with genetic correlation analysis [36]. It should also be noted that, while pleiotropy can give rise to genetic correlation, other factors could also influence an observed correlation, including misclassification of individuals into either disease group [115]. The SLI Consortium sample was examined for autism, and samples were excluded if they had an indication of autism; similarly, the ICD criteria require specific social and behavioral impairments for a diagnosis of childhood autism, which a child with SLI should not typically exhibit. However, it cannot be ruled out that a misclassification did occur.

Conclusions

Our study did not find significant genetic overlaps between SLI and ASD, ADHD, and schizophrenia. However, a small but significant genetic overlap between SLI and childhood autism, in particular, was found. As this

was not observed for Asperger's syndrome, and the difference in PRS between the two case groups was significant, it may suggest that these two disorders, which differ linguistically, could also be distinguished genetically using polygenic risk scores for language impairment. While we found some overlaps across candidate genes for SLI and ASD, childhood autism, and schizophrenia, these associations did not survive Bonferroni correction and can, at most, provide suggestive evidence for pleiotropy. Taken together, our results may suggest that there is a number of loci that influence "pre-linguistic" mechanisms that influence neurodevelopment in general and thus may have an impact on language ability down the road, or loci that influence linguistic ability that is not domain-specific, which are shared between SLI and ASD, and, in particular, SLI and childhood autism. However, at this point, this is only speculative. Larger discovery samples may be needed in order to obtain a more reliable PRS, and larger exome-sequenced samples may be needed to detect the effects of rare variants, although our results may also suggest that rare variants in language-related genes may not have pleiotropic effects on the investigated neurodevelopmental disorders.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s11689-021-09403-z.

Additional file 1. Supplementary notes. Furtherinformation about the EMIM models.

Additional file 2: Supplementary Figure S1. Flowchart for the rare variant analyses.

Additional file 3: Supplementary Table S1. Top 10 SNPs from the summary statistics from the SLI GWAS.

Authors' contributions
RN conceived and designed the study, performed the QC steps for the final genetic dataset as used in this study, performed the genetic and statistical analyses, analyzed the results, and wrote the manuscript; VA performed the

Nudel *et al. Journal of Neurodevelopmental Disorders*   (2021) 13:54

Page 12 of 15

first steps of the quality control of the genetic markers from the genotyping array prior to sample filtering and assisted with the preparation of the exome-sequencing input data; AB made intellectual contributions to the study design; MN and TW are principal investigators in the iPSYCH Consortium who supervised the pilot study (MN) and the extended study (TW). All authors have read and approved the final manuscript.

Availability of data and materials
The analyses performed in this study used summary statistics and datasets from previously published papers, as referenced in the text; no discovery analyses were performed. Data availability information for these datasets can be found in the relevant papers.

Declarations

Ethics approval and consent to participate
The Danish Scientific Ethics Committee approved this study (ESDH 1-10-72-287-12). The following institutions also approved the study: the Danish Health Data Authority, the Danish data protection agency and the Danish Neonatal Screening Biobank Steering Committee. All personal information from the registers is anonymized when used for research purposes, according to Danish legislation; informed consent from participants was not required.

Consent for publication
Not applicable.

Competing interests
All researchers had full independence from the funders. The authors report no biomedical financial interests or potential conflicts of interest. TW states that he has acted as a lecturer and scientific counselor to H. Lundbeck A/S.

Author details
[1]Institute of Biological Psychiatry, Mental Health Centre Sct. Hans, Mental Health Services Copenhagen, Roskilde, Denmark. [2]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Aarhus, Denmark. [3]CORE - Copenhagen Research Centre for Mental Health, Mental Health Centre Copenhagen, Copenhagen University Hospital, Copenhagen, Denmark. [4]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.

References
1. Lightbown PM, White L. The influence of linguistic theories on language acquisition research: description and explanation. Lang Learn. 1987;37(4):483–510.
2. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015;47(11):1236–41 Epub 29 Sep 2015.
3. Stromswold K. The heritability of language: a review and metaanalysis of twin, adoption, and linkage studies. Language. 2001;77(4):647–723.
4. Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. A forkhead-domain gene is mutated in a severe speech and language disorder. Nature. 2001;413(6855):519–23 Epub 5 Oct 2001.
5. Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet. 2013;45(9):984–94 Epub 13 Aug 2013.
6. Dover CJ, Le Couteur A. How to diagnose autism. Arch Dis Child. 2007;92(6):540–5 Epub 23 May 2007.
7. Baird J, Stevenson JC, Williams DC. The evolution of ADHD: a disorder of communication? Q Rev Biol. 2000;75(1):17–35 Epub 18 March 2000.
8. Bishop DVM. What causes specific language impairment in children? Curr Dir Psychol Sci. 2006;15(5):217–21 Epub 15 Nov 2008.
9. Bishop DVM, Snowling MJ, Thompson PA, Greenhalgh T. Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: terminology. J Child Psychol Psychiatry Allied Discip. 2017;58(10):1068–80 Epub 04 April 2017.
10. Newbury DF, Bishop DVM, Monaco AP. Genetic influences on language impairment and phonological short-term memory. Trends Cogn Sci. 2005;9(11):528–34 Epub 29 Sep 2005.
11. Newbury DF, Bonora E, Lamb JA, Fisher SE, Lai CS, Baird G, et al. FOXP2 is not a major susceptibility gene for autism or specific language impairment. Am J Hum Genet. 2002;70(5):1318–27 Epub 15 March 2002.
12. DeLisi LE. Speech disorder in schizophrenia: review of the literature and exploration of its relation to the uniquely human capacity for language. Schizophr Bull. 2001;27(3):481–96 Epub 13 Oct 2001ss.
13. Levy DL, Coleman MJ, Sung H, Ji F, Matthysse S, Mendell NR, et al. The genetic basis of thought disorder and language and communication disturbances in schizophrenia. J Neurolinguistics. 2010;23(3):176 Epub 18 Feb 2010.
14. Li T, Zeng Z, Zhao Q, Wang T, Huang K, Li J, et al. FoxP2 is significantly associated with schizophrenia and major depression in the Chinese Han population. World J Biol Psychiatry. 2013;14(2):146–50 Epub 13 March 2012.
15. Tolosa A, Sanjuan J, Dagnall AM, Molto MD, Herrero N, de Frutos R. FOXP2 gene and language impairment in schizophrenia: association and epigenetic studies. BMC Med Genet. 2010;11:114 Epub 24 July 2010.
16. Yin J, Jia N, Liu Y, Jin C, Zhang F, Yu S, et al. No association between FOXP2 rs10447760 and schizophrenia in a replication study of the Chinese Han population. Psychiatr Genet. 2018;28(2):19–23 Epub 19 Jan 2018.
17. McCarthy NS, Clark ML, Jablensky A, Badcock JC. No association between common genetic variation in FOXP2 and language impairment in schizophrenia. Psychiatry Res. 2019;271:590–7 Epub 17 Dec 2018.
18. Schork AJ, Won H, Appadurai V, Nudel R, Gandal M, Delaneau O, et al. A genome-wide association study of shared risk across psychiatric disorders implicates gene regulation during fetal neurodevelopment. Nat Neurosci. 2019;22(3):353–61 Epub 30 Jan 2019.
19. Bishop DVM. Autism and specific language impairment: categorical distinction or continuum? Novartis Found Symp. 2003;251:213–26 discussion 26-34, 81-97.
20. Kjelgaard MM, Tager-Flusberg H. An investigation of language impairment in autism: implications for genetic subgroups. Lang Cogn Process. 2001;16(2-3):287–308 Epub 17 May 2006.
21. Whitehouse AJ, Barry JG, Bishop DVM. The broader language phenotype of autism: a comparison with specific language impairment. J Child Psychol Psychiatry Allied Discip. 2007;48(8):822–30 Epub 09 Aug 2007.
22. Williams D, Payne H, Marshall C. Non-word repetition impairment in autism and specific language impairment: evidence for distinct underlying cognitive causes. J Autism Dev Disord. 2013;43(2):404–17 Epub 27 June 2012.
23. Leyfer OT, Tager-Flusberg H, Dowd M, Tomblin JB, Folstein SE. Overlap between autism and specific language impairment: comparison of Autism Diagnostic Interview and Autism Diagnostic Observation Schedule scores. Autism Res. 2008;1(5):284–96 Epub 11 April 2009.
24. Williams D, Botting N, Boucher J. Language in autism and specific language impairment: where are the links? Psychol Bull. 2008;134(6):944–63 Epub 29 Oct 2008.
25. Vernes SC, Newbury DF, Abrahams BS, Winchester L, Nicod J, Groszer M, et al. A functional genetic link between distinct developmental language disorders. N Engl J Med. 2008;359(22):2337–45 Epub 07 Nov 2008.

Nudel *et al. Journal of Neurodevelopmental Disorders*   (2021) 13:54

Page 13 of 15

26. Alarcon M, Abrahams BS, Stone JL, Duvall JA, Perederiy JV, Bomar JM, et al. Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. Am J Hum Genet. 2008;82(1):150–9 Epub 03 Jan 2008.

27. Friedman JI, Vrijenhoek T, Markx S, Janssen IM, van der Vliet WA, Faas BH, et al. CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. Mol Psychiatry. 2008;13(3):261–6.

28. 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 28 Feb 2014.

29. Zuko A, Kleijer KT, Oguro-Ando A, Kas MJ, van Daalen E, van der Zwaag B, et al. Contactins in the neurobiology of autism. Eur J Pharmacol. 2013;719(1-3):63–74 Epub 23 July 2013.

30. Weyn-Vanhentenryck SM, Mele A, Yan Q, Sun S, Farny N, Zhang Z, et al. HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. Cell Rep. 2014;6(6):1139–52 Epub 13 March 2014.

31. Kalikiri MK, Mamidala MP, Rao AN, Rajesh V. Analysis and functional characterization of sequence variations in ligand binding domain of thyroid hormone receptors in autism spectrum disorder (ASD) patients. Autism Res. 2017;10(12):1919–28 Epub 01 Sep 2017.

32. Meehan TF, Carr CJ, Jay JJ, Bult CJ, Chesler EJ, Blake JA. Autism candidate genes via mouse phenomics. J Biomed Inform. 2011;44(Suppl 1):S5–11 Epub 15 March 2011.

33. Bishop DVM. Overlaps between autism and language impairment: phenomimicry or shared etiology? Behav Genet. 2010;40(5):618–29 Epub 20 July 2010.

34. 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 17 May 2011.

35. Wray NR, Wijmenga C, Sullivan PF, Yang J, Visscher PM. Common disease is more complex than implied by the core gene omnigenic model. Cell. 2018;173(7):1573–80 Epub 16 June 2018.

36. Wray NR, Lee SH, Mehta D, Vinkhuyzen AA, Dudbridge F, Middeldorp CM. Research review: polygenic methods and their application to psychiatric traits. J Child Psychol Psychiatry Allied Discip. 2014;55(10):1068–87 Epub 19 Aug 2014.

37. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009;460(7256):748–52 Epub 03 July 2009.

38. Belsky DW, Israel S. Integrating genetics and social science: genetic risk scores. Biodemography Soc Biol. 2014;60(2):137–55.

39. Fullerton J, Nurnberger J. Polygenic risk scores in psychiatry: will they be useful for clinicians? [version 1; peer review: 4 approved]. F1000Research. 2019;8(1293). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6676506/.

40. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. Nat Rev Genet. 2018;19(9):581–90 Epub 24 May 2018.

41. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018;50(9):1219–24 Epub 15 Aug 2018.

42. Hamshere ML, O'Donovan MC, Jones IR, Jones L, Kirov G, Green EK, et al. Polygenic dissection of the bipolar phenotype. Br J Psychiatry. 2011;198(4):284–8 Epub 06 Oct 2011.

43. Landi I, Kaji D, Cotter L, Vleck TV, Belbin G, Preuss M, et al. Polygenic risk scores lack prognostic value for adults with severe mental illness. medRxiv. 2021:2021.03.19.21253906. https://www.medrxiv.org/content/10.1101/2021.03.19.21253906v1.

44. Bishop DVM. Why is it so hard to reach agreement on terminology? The case of developmental language disorder (DLD). Int J Language Commun Disord. 2017;52(6):671–80.

45. Newbury DF, Gibson JL, Conti-Ramsden G, Pickles A, Durkin K, Toseeb U. Using polygenic profiles to predict variation in language and psychosocial outcomes in early and middle childhood. J Speech Language Hear Res. 2019;62(9):3381–96.

46. Dale PS, Stumm S, Selzam S, Hayiou-Thomas ME. Does the inclusion of a genome-wide polygenic score improve early risk prediction for later language and literacy delay? J Speech Language Hear Res. 2020;63(5):1467–78.

47. Kimball EE, Sayce L. Research in speech science and voice disorders: the promise of modern genetic approaches in improving clinical diagnosis and treatment. Perspect ASHA Special Interest Groups. 2020;5(6):1828–38.

48. Gialluisi A, Andlauer TFM, Mirza-Schreiber N, Moll K, Becker J, Hoffmann P, et al. Genome-wide association scan identifies new variants associated with a cognitive predictor of dyslexia. Transl Psychiatry. 2019;9(1):77 Epub 12 Feb 2019.

49. 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 03 Oct 2019.

50. Pedersen CB, Bybjerg-Grauholm J, Pedersen MG, Grove J, Agerbo E, Baekvad-Hansen M, et al. The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. Mol Psychiatry. 2017;23:6–14 Epub 20 Sep 2017.

51. Nudel R, Appadurai V, Schork AJ, Buil A, Bybjerg-Grauholm J, Borglum AD, et al. A large population-based investigation into the genetics of susceptibility to gastrointestinal infections and the link between gastrointestinal infections and mental illness. Hum Genet. 2020;139(5):593–604 Epub 11 March 2020.

52. Nudel R, Benros ME, Krebs MD, Allesoe RL, Lemvigh CK, Bybjerg-Grauholm J, et al. Immunity and mental illness: findings from a Danish population-based immunogenetic study of seven psychiatric and neurodevelopmental disorders. Eur J Hum Genet. 2019;27(9):1445–55 Epub 13 April 2019.

53. Nudel R, Wang Y, Appadurai V, Schork AJ, Buil A, Agerbo E, et al. A large-scale genomic investigation of susceptibility to infection and its association with mental disorders in the Danish population. Transl Psychiatry. 2019;9(1):283 Epub 13 Nov 2019.

54. World Health Organization. WHO ICD-10: Psykiske Lidelser Og Adfærdsmæssige Forstyrelser. Klassifikation Og Diagnosekriterier [WHO ICD-10: Mental and Behavioural Disorders. Classification and Diagnostic Criteria]. Copenhagen: World Health Organization; 1994.

55. World Health Organization. Klassifikation Af Sygdomme; Udvidet Dansk-Latinsk Udgave Af Verdenssundhedsorganisationens Internationale Klassifikation Af Sygdomme. 8 Revision, 1965 [Classification of Diseases: Extended Danish-Latin Version of the World Health Organization International Classification of Diseases]. Copenhagen; 1971.

56. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;51(3):431–44 Epub 23 Feb 2019.

57. 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.

58. Erlangsen A, Appadurai V, Wang Y, Turecki G, Mors O, Werge T, et al. Genetics of suicide attempts in individuals with and without mental disorders: a population-based genome-wide association study. Mol Psychiatry. 2020;25(10):2410–21 Epub 18 Aug 2018.

59. Falcaro M, Pickles A, Newbury DF, Addis L, Banfield E, Fisher SE, et al. Genetic and phenotypic effects of phonological short-term memory and grammatical morphology in specific language impairment. Genes Brain Behav. 2008;7(4):393–402 Epub 17 Nov 2007.

60. The SLI Consortium. Highly significant linkage to the SLI1 locus in an expanded sample of individuals affected by specific language impairment. Am J Hum Genet. 2004;74(6):1225–38 Epub 11 May 2004.

61. The SLI Consortium. A genomewide scan identifies two novel loci involved in specific language impairment. Am J Hum Genet. 2002;70(2):384–98 Epub 16 Jan 2002.

62. Howey R, Cordell HJ. PREMIM and EMIM: tools for estimation of maternal, imprinting and interaction effects using multinomial modelling. BMC Bioinform. 2012;13:149 Epub 29 June 2012.

63. 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 25 Dec 2010.

64. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. GigaScience. 2019;8(7):giz082.

65. Dudbridge F. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 2013;9(3):e1003348 Epub 05 April 2013.

66. Ware EB, Schmitz LL, Faul JD, Gard A, Mitchell C, Smith JA, et al. Heterogeneity in polygenic scores for common human traits. bioRxiv. 2017:106062. https://pubmed.ncbi.nlm.nih.gov/31307061/.

67. Lee SH, Goddard ME, Wray NR, Visscher PM. A better coefficient of determination for genetic profile analysis. Genet Epidemiol. 2012;36(3):214–24 Epub 21 June 2012.

68. Baird G, Simonoff E, Pickles A, Chandler S, Loucas T, Meldrum D, et al. Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). Lancet. 2006;368(9531):210–5 Epub 18 July 2006.

69. Polanczyk G, de Lima MS, Horta BL, Biederman J, Rohde LA. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. Am J Psychiatry. 2007;164(6):942–8 Epub 02 June 2007.

70. Bhugra D. The global prevalence of schizophrenia. PLoS Med. 2005;2(5):e151 quiz e75. Epub 27 May 2005.

71. Mattila ML, Kielinen M, Jussila K, Linna SL, Bloigu R, Ebeling H, et al. An epidemiological and diagnostic study of Asperger syndrome according to four sets of diagnostic criteria. J Am Acad Child Adolesc Psychiatry. 2007;46(5):636–46 Epub 24 April 2007.

72. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2014.

73. Chen XS, Reader RH, Hoischen A, Veltman JA, Simpson NH, Francks C, et al. Next-generation DNA sequencing identifies novel gene variants and pathways involved in specific language impairment. Sci Rep. 2017;7:46105 Epub 26 April 2017.

74. Mountford HS, Newbury DF. The genomic landscape of language: insights into evolution. J Language Evol. 2017;3(1):49–58.

75. Guerra J, Cacabelos R. Genomics of speech and language disorders. J Transl Genet Genomics. 2019;3:9.

76. 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 09 July 2020.

77. Newbury DF, Winchester L, Addis L, Paracchini S, Buckingham LL, Clark A, et al. CMIP and ATP2C2 modulate phonological short-term memory in language impairment. Am J Hum Genet. 2009;85(2):264–72 Epub 04 Aug 2009.

78. Villanueva P, Nudel R, Hoischen A, Fernandez MA, Simpson NH, Gilissen C, et al. Exome sequencing in an admixed isolated population indicates NFXL1 variants confer a risk for specific language impairment. PLoS Genet. 2015;11(3):e1004925 Epub 18 March 2015.

79. Kornilov SA, Rakhlin N, Koposov R, Lee M, Yrigollen C, Caglayan AO, et al. Genome-wide association and exome sequencing study of language disorder in an isolated population. Pediatrics. 2016;137(4):e20152469. https://pubmed.ncbi.nlm.nih.gov/27016271/.

80. Eicher JD, Powers NR, Miller LL, Akshoomoff N, Amaral DG, Bloss CS, et al. Genome-wide association study of shared components of reading disability and language impairment. Genes Brain Behav. 2013;12(8):792–801 Epub 13 Sep 2013.

81. Luciano M, Evans DM, Hansell NK, Medland SE, Montgomery GW, Martin NG, et al. A genome-wide association study for reading and language abilities in two population cohorts. Genes Brain Behav. 2013;12(6):645–52.

82. Ersland KM, Christoforou A, Stansberg C, Espeseth T, Mattheisen M, Mattingsdal M, et al. Gene-based analysis of regionally enriched cortical genes in GWAS data sets of cognitive traits and psychiatric disorders. PLoS One. 2012;7(2):e31687 Epub 03 March 2012.

83. 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 Neurodev Disord. 2014;6(1):1 Epub 18 Oct 2014.

84. Torres AR, Maciulis A, Stubbs EG, Cutler A, Odell D. The transmission disequilibrium test suggests that HLA-DR4 and DR13 are linked to autism spectrum disorder. Hum Immunol. 2002;63(4):311–6 Epub 01 June 2002.

85. Odell JD, Warren RP, Warren WL, Burger RA, Maciulis A. Association of genes within the major histocompatibility complex with attention deficit hyperactivity disorder. Neuropsychobiology. 1997;35(4):181–6 Epub 01 Jan 1997.

86. Warren RP, Odell JD, Warren WL, Burger RA, Maciulis A, Daniels WW, et al. Strong association of the third hypervariable region of HLA-DR beta 1 with autism. J Neuroimmunol. 1996;67(2):97–102 Epub 01 July 1996.

87. Wright P, Nimgaonkar VL, Donaldson PT, Murray RM. Schizophrenia and HLA: a review. Schizophr Res. 2001;47(1):1–12 Epub 13 Feb 2001.

88. Satterstrom FK, Walters RK, Singh T, Wigdor EM, Lescai F, Demontis D, et al. Autism spectrum disorder and attention deficit hyperactivity disorder have a similar burden of rare protein-truncating variants. Nat Neurosci. 2019;22(12):1961–5 Epub 27 Nov 2019.

89. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27(21):2987–93 Epub 10 Sep 2011.

90. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly. 2012;6(2):80–92 Epub 26 June 2012.

91. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet. 2012;91(2):224–37 Epub 07 Aug 2012.

92. Kang HM. EPACTS (Efficient and Parallelizable Association Container Toolbox). v3.2.6 ed; 2014.

93. 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 04 Oct 2015.

94. Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith E, O'Brien M. Prevalence of specific language impairment in kindergarten children. J Speech Language Hear Res. 1997;40(6):1245–60.

95. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinform. 2011;12:77 Epub 19 March 2011.

96. Sanders JL. Qualitative or quantitative differences between Asperger's disorder and autism? Historical considerations. J Autism Dev Disord. 2009;39(11):1560–7 Epub 24 June 2009.

97. Koning C, Magill-Evans J. Social and language skills in adolescent boys with Asperger syndrome. Autism. 2001;5(1):23–36 Epub 16 Nov 2001.

98. Saalasti S, Lepisto T, Toppila E, Kujala T, Laakso M, Nieminen-von Wendt T, et al. Language abilities of children with Asperger syndrome. J Autism Dev Disord. 2008;38(8):1574–80 Epub 08 March 2008.

99. Aikawa J, Grobe K, Tsujimoto M, Esko JD. Multiple isozymes of heparan sulfate/heparin GlcNAc N-deacetylase/GlcN N-sulfotransferase. Structure and activity of the fourth member, NDST4. J Biol Chem. 2001;276(8):5876–82 Epub 23 Nov 2000.

100. Pan Y, Luo X, Liu X, Wu LY, Zhang Q, Wang L, et al. Genome-wide association studies of maximum number of drinks. J Psychiatr Res. 2013;47(11):1717–24 Epub 21 Aug 2013.

101. Qi Q, Menzaghi C, Smith S, Liang L, de Rekeneire N, Garcia ME, et al. Genome-wide association analysis identifies TYW3/CRYZ and NDST4 loci associated with circulating resistin levels. Hum Mol Genet. 2012;21(21):4774–80 Epub 31 July 2012.

102. Singh T, Neale BM, Daly MJ, Schizophrenia Exome Meta-Analysis (SCHEMA) Consortium. Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia. medRxiv. 2020:2020.09.18.20192815.

103. Gawlas K, Stunnenberg HG. Differential binding and transcriptional behaviour of two highly related orphan receptors, ROR alpha(4) and ROR beta(1). Biochim Biophys Acta. 2000;1494(3):236–41 Epub 21 Dec 2000.

104. McGrath CL, Glatt SJ, Sklar P, Le-Niculescu H, Kuczenski R, Doyle AE, et al. Evidence for genetic association of RORB with bipolar disorder. BMC Psychiatry. 2009;9:70 Epub 17 Nov 2009.

105. Smith V, Mirenda P, Zaidman-Zait A. Predictors of expressive vocabulary growth in children with autism. J Speech Language Hear Res. 2007;50(1):149–60 Epub 09 March 2007.

106. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An J-Y, et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. Cell. 2020;180(3):568–84 e23.
107. Piazza R, Magistroni V, Redaelli S, Mauri M, Massimino L, Sessa A, et al. SETBP1 induces transcription of a network of development genes by acting as an epigenetic hub. Nat Commun. 2018;9(1):2192 Epub 08 June 2018.
108. Eising E, Carrion-Castillo A, Vino A, Strand EA, Jakielski KJ, Scerri TS, et al. A set of regulatory genes co-expressed in embryonic human brain is implicated in disrupted speech development. Mol Psychiatry. 2019;24(7):1065–78 Epub 22 Feb 2018.
109. Filges I, Shimojima K, Okamoto N, Rothlisberger B, Weber P, Huber AR, et al. Reduced expression by SETBP1 haploinsufficiency causes developmental and expressive language delay indicating a phenotype distinct from Schinzel-Giedion syndrome. J Med Genet. 2011;48(2):117–22 Epub 03 Nov 2010.
110. Marseglia G, Scordo MR, Pescucci C, Nannetti G, Biagini E, Scandurra V, et al. 372 kb microdeletion in 18q12.3 causing SETBP1 haploinsufficiency associated with mild mental retardation and expressive speech impairment. Eur J Med Genet. 2012;55(3):216–21 Epub 16 Feb 2012.
111. Coe BP, Witherspoon K, Rosenfeld JA, van Bon BW, Vulto-van Silfhout AT, Bosco P, et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat Genet. 2014;46(10):1063–71 Epub 15 Sep 2014.
112. Lencz T, Yu J, Kahn RR, Carmi S, Lam M, Ben-Avraham D, et al. Novel ultra-rare exonic variants identified in a founder population implicate cadherins in schizophrenia. Neuron. 2021; 109(9):1465–78.e4. Epub 2021 Mar 22.
113. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511(7510):421–7 Epub 25 July 2014.
114. Bellou E, Stevenson-Hoare J, Escott-Price V. Polygenic risk and pleiotropy in neurodegenerative diseases. Neurobiol Dis. 2020;142:104953 Epub 24 May 2020.
115. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. Nat Rev Genet. 2013;14(7):483–95 Epub 12 June 2013.

Publisher's Note