Lingo: an automated, web-based deep phenotyping platform for language ability

Lucas G. Casten[1,2], Tanner Koomar[2], Muhammad Elsadany[1,2], Caleb McKone[2], Ben Tysseling[2], Mahesh Sasidharan[2], J. Bruce Tomblin[3], Jacob J. Michaelson[2,3,4,5*]

1 Interdisciplinary Graduate Program in Genetics, University of Iowa, Iowa City, IA
2 Department of Psychiatry, University of Iowa, Iowa City, IA
3 Department of Communication Sciences and Disorders, University of Iowa, Iowa City, IA
4 Iowa Neuroscience Institute, University of Iowa, Iowa City, IA
5 Hawkeye Intellectual and Developmental Disabilities Research Center (Hawk-IDDRC), University of Iowa, Iowa City, IA
***** jacob-michaelson@uiowa.edu

Abstract

Background:
Language and the ability to communicate effectively are key factors in mental health and well-being. Despite this critical importance, research on language is limited by the lack of a scalable phenotyping toolkit.

Methods:
Here, we describe and showcase Lingo – a flexible online battery of language and nonverbal reasoning skills based on seven widely used tasks (COWAT, picture narration, vocal rhythm entrainment, rapid automatized naming, following directions, sentence repetition, and nonverbal reasoning). The current version of Lingo takes approximately 30 minutes to complete, is entirely open source, and allows for a wide variety of performance metrics to be extracted. We asked > 1,300 individuals from multiple samples to complete Lingo, then investigated the validity and utility of the resulting data.

Results:
We conducted an exploratory factor analysis across 14 features derived from the seven assessments, identifying five factors. Four of the five factors showed acceptable test-retest reliability (Pearson's R > 0.7). Factor 2 showed the highest reliability (Pearson's R = 0.95) and loaded primarily on sentence repetition task performance. We validated Lingo with objective measures of language ability by comparing performance to gold-standard assessments: CELF-5 and the VABS-3. Factor 2 was significantly associated with the CELF-5 "core language ability" scale (Pearson's R = 0.77, p-value < 0.05) and the VABS-3 "communication" scale (Pearson's R = 0.74, p-value < 0.05). Factor 2 was positively associated with phenotypic and genetic measures of socieconomic status. Interestingly, we found the parents of children with language impairments had lower Factor 2 scores (p-value < 0.01). Finally, we found Lingo factor scores were significantly predictive of numerous psychiatric and neurodevelopmental conditions.

Conclusions:
Together, these analyses support Lingo as a powerful platform for scalable deep phenotyping of language and other cognitive abilities. Additionally, exploratory analyses provide supporting evidence for the heritability of language ability and the complex relationship between mental health and language.

1 Introduction

Language is a complex phenomenon that relies on multiple cognitive domains and brain networks [1, 2]. Across individuals there is incredible variability in language ability and speech patterns [3]. Studies of language ability, particularly specific language impairments, have shown language ability is a highly genetic trait (twin based heritability estimates range from 45-60%) [4, 5]. Despite evidence for substantial heritability, the relatively low power of candidate gene association studies has led to a small handful of tenuously implicated genes and genetic loci [6]. Successful identification of these language-associated genes depends on our ability to uniformly phenotype individuals across the domains of language, and at the large scale needed for genetic research. Traditionally, gold standard assessment of language ability has depended on paper-and-pencil, in-person proctored tests. These assessments are expensive to administer, time consuming, and require highly trained proctors. Beyond the challenges related to collecting a large enough sample for a genome wide association study (GWAS) with this method, some of these tests

are not amenable to genetic analysis in the general population because they are designed to detect clinically relevant language impairment. This may result in a low ceiling, thus limiting their usefulness in assessing inter-individual variation in language ability in population samples. While recent efforts from the GenLang consortium have shown significant GWAS associations with traditional language assessments in a large meta-analysis of more than 30,000 individuals, they also discussed the difficulty of collecting additional data with these assessments and the need for a scalable deep language phenotyping tool to facilitate future research [7]. Recent work has shown short online assessments can accurately measure general intelligence that has a genetic basis [8]. Unfortunately, current online assessments lack the flexibility to capture a variety of speech and language related phenotypes. Development of a language and cognition toolkit that allows for a variety of data to be collected, like audio and selection data, would be incredibly valuable for deep phenotyping in future genetic research.

Language is an important endophenotype in psychiatric research. Children with language impairments are more likely to develop psychiatric disorders later in life [9]. Many individuals with psychiatric and neurodevelopmental disorders show aberrant speech patterns. Schizophrenia is associated with deficits in receptive and expressive language as well as decreased coherence between sentences when describing pictures [10, 11]. Major depression has been linked to less frequent use of words with positive meaning and more monotone speech [12]. Autism, a neurodevelopmental condition, is highly comorbid with language impairments. It is estimated that 25-30% of autistic children at the age of 5 do not speak at all or use very few words when speaking [13]. Collecting rich speech data from populations with psychiatric and neurodevelopmental conditions may lead to more efficient diagnosis and monitoring [14].

To begin to remove the methodological barriers to study language, we have developed a battery of language and related cognitive tasks that are administered remotely and unsupervised through a web browser. This web-based language battery records a variety of user-generated data, including item selections (with time-to-selection) and recorded audio, from which several latent phenotypes can be derived. Here, we describe this battery, the rationale for selecting the tasks involved, and several validations of the resulting data across multiple samples. We then conduct exploratory analysis of the derived phenotypes and show significant associations with mental health, genetic risk, and socioeconomic status. Together, these results suggest that this web-based battery is a promising tool for quantitatively phenotyping multiple domains of language and cognitive abilities.

2 Results

2.1 Lingo overview

An overview of the Lingo battery can be seen in Figure 1. Prompt images can be seen in the Supplementary Figures and descriptions of each task can be found in the "Methods and materials section". Code is available at https://research-git.uiowa.edu/michaelson-lab/language-screener/ScreenerVersion2/-/tree/master.

2.2 Lingo performance features and their factor structure

The open-ended nature of many tasks in the Lingo screener offers flexibility for measuring task performance, allowing for both hypothesis-driven and exploratory data analysis. Throughout this manuscript we showcase the utility of 14 features from the Lingo tasks in measuring different aspects of language ability. All derived features used in our analyses are listed in Figure 2A and described in table 2. In brief, we extracted performance metrics related to concepts like verbal memory, verbal fluency, vocabulary, reading speed, receptive language ability, rhythm, and nonverbal ability.

Due to the high intercorrelation of these features, we conducted an exploratory factor analysis across a sample of N = 1,344 unique individuals from multiple samples (descriptions can be found in 1. First, we found the Lingo data is factorable – as measured by Bartlett's Test of Sphericity (p-value $<$ 2e-16) and the Kaiser-Meyer-Olkin criterion test on the correlation matrix of the Lingo variables (KMO value $>$ 0.6). Next, we identified five distinct factors from our 14 input Lingo features using a data-driven parallel analysis approach (see Figure 2A). Factor 1 ("poor verbal fluency") loaded negatively onto number of unique words said during the COWAT and positively onto reading time. Factor 2 ("core language") loaded strongly sentence repetition task performance as well as weaker loading from the following directions and matrix reasoning tasks. Factor 3 ("rhythm/timkeeping") loaded mostly onto performance metrics from the sync voice task. Factor 4 ("talkativeness/descriptiveness") loaded onto number of words said during the picture narration task. Finally, Factor 5 ("vocabulary") loaded primarily onto our vocabulary measure and hapax count, suggesting individuals scoring high on this factor frequently used word others did not. Factor score inter-correlations are shown in Figure 2B. Of note, Factors 1 and 5 had the strongest correlation (Pearson's R = -0.56, p-value $<$ 0.01).

2.3 Test-retest reliability

To investigate test-retest reliability of Lingo performance, we asked 14 independent adults from the SPARK research match study to take the Lingo battery twice (separated by at least two weeks between tests). Test-retest reliability was measured by calculating Pearson's correlations between the factor score for the first and second testing sessions, on a factor-wise basis (Figure 2C). Factors 1, 2, 4, and 5 showed acceptable test-retest reliability (Pearson's R > 0.7). Factor 2 showed particularly high test-retest reliability (Pearson's R = 0.95, p-value < 0.01). Factor 3 (rhythm/vocal timekeeping) did not show strong reliability (Pearson's R = 0.46, p-value = 0.1) These results suggest that many of the factors derived from Lingo data are reliable, in particular the verbal memory related variables.

2.4 Agreement with a gold-standard language assessments

To establish concurrent validity between the Lingo battery and well-established measures of language ability, we compared our factor scores to gold-standard clinical assessments: the CELF-5 and the Vineland Adaptive Behavior Scales 3 (VABS-3) [15, 16]. For the CELF-5 test validity analysis we recruited 8 individuals from a local neurodevelopmental registry (devGenes, Table 1) who completed Lingo and were assessed with the CELF-5 on the same day. Factor scores from Lingo were then correlated with their age-normed CELF-5 clinical subscale scores (Figure 3). Factor 2 had a significant correlation with the "core language score percentile" from the CELF-5 (Pearson's R = 0.77, p-value = 0.03, Figure 3A). Next, we conducted a similar analysis in our SPARK sample. Eight individuals in our SPARK sample also had complete VABS-3 data. We found that factor 2 was significantly associated with the VABS-3 "communication" standard score (Pearson's R = 0.74, p-value = 0.04, Figure 3B).

2.5 Socioeconomic correlations with Lingo factors

Given the findings that Lingo scores are reliable and valid, we sought to showcase the utility of our Lingo factors in a series of phenotypic and genetic analyses. While previous research has found robust links between socioeconomic status (SES) and language test scores in children [17–19], less is known about how SES influences specific domains of language later in life. Common measurements of SES include educational attainment (highest level of education completed) and the Area Deprivation Index (ADI – a normed measure of socioeconomic disadvantage based on quality of life variables mapped to the zip code of someone's residence). To determine the relationship of SES with different language and cognitive scores in adults, we investigated possible associations between our Lingo factor scores with educational attainment and ADI. Lingo factor 2 showed a significant positive correlation with educational attainment in adults at least 30 years old (Pearson's R = 0.28, p-value < 0.01, N = 250, Figure 3C). Similarly, we found ADI was negatively correlated with factor 2 scores (Pearson's R = -0.08, p-value < 0.01, N = 1243, Figure 3D). The associations between SES and Lingo scores were persistent. All five Lingo factors were significantly correlated with education level and ADI.

2.6 Polygenic score associations with Lingo factors

In the previous section we describe evidence linking phenotypic measures of SES to Lingo factors. To investigate whether these findings had support at the biological level, we tested for associations with genetic estimates of SES. We computed 12 different polygenic scores (PGS) for 811 individuals of European descent in our SPARK sample with genotype data to link with our Lingo factors. The 12 PGS we calculated were relevant to cognitive ability, SES, and neuropsychiatric risk. All PGS were corrected for population stratification prior to analysis. We found factor 2 was significantly correlated with educational attainment PGS (Pearson R = 0.20, FDR adjusted p-value < 0.01) and cognitive performance PGS (Pearson R = 0.20, FDR adjusted p-value < 0.01, Figure 4A). Factor 2 also showed a nominally significant association with depression PGS (Pearson R = -0.08, uncorrected p-value = 0.02) Interestingly, factor 4 (talkativeness/descriptiveness) was associated with decreased polygenic risk for schizophrenia (Pearson R = -0.11, FDR adjusted p-value = 0.03) and depression (Pearson R = -0.10, FDR adjusted p-value = 0.04).

2.7 Familiality of language ability

Using our SPARK sample, we investigated whether our factor scores derived from Lingo data showed evidence of familiality, which would support the well-established heritability of language ability [4, 5]. Upon enrolling in SPARK, parents are asked to characterize their autistic child's language ability with Likert-scale responses ranging from: "Uses longer sentences of his/her own and is able to tell you something that happened" to "No words/does not speak". This Likert-scale allowed us to binarize their child's language ability into an unimpaired language group (N = 693, "Uses longer sentences of his/her own and is able to tell you something that happened"), and an impaired language group

(N = 310, "Combines 3 words together into short sentences" or "Uses single words meaningfully" or "No words/does not speak"). Indeed, we found that parents' Factor 2 score was significantly associated with their children's language ability (t-statistic = 3.7, p-value < 0.01, Figure 4B) – suggesting parents of autistic children with language impairments have poorer verbal memory themselves. Notably, parents' Factor 2 score was independently predictive of their child's language impairment after controlling for the child's cognitive impairment, age, and sex (beta = -0.16, Z = -2.05, p-value = 0.04, Figure 4C). These results support the heritability of language ability that has been previously reported, and further underscore the validity of the data yielded through Lingo.

2.8 Lingo factors are predictive of psychiatric diagnoses

Finally, we tested whether Lingo captures language and cognitive phenotypic variation relevant to psychiatric diagnosis. A subset of SPARK participants (N = 845) completed an additional self-report questionnaire about psychiatric, neurodevelopmental, or neurological diagnoses they have received. We then fit logistic regression models to predict diagnosis from Lingo factor scores while controlling for age and sex of the participant. Lingo factor 2 was negatively associated with endorsement of language impairment (beta = -0.6, Z = -4.1, FDR adjusted p-value < 0.01), intellectual disability (beta = -0.5, Z = -5, FDR adjusted p-value < 0.01), and bipolar disorder (beta = -0.29, Z = -2.84, FDR adjusted p-value = 0.04), Figure 5. Factor 5 was associated with higher endorsement of autism (beta = 0.51, Z = 5.1, FDR adjusted p-value < 0.01, Figure 5).

3 Discussion

Lingo is a powerful platform for scalable deep phenotyping of language ability. With multiple samples collected from a variety of settings we were able to validate the utility of Lingo in quantifying distinct language domains. We identified a factor related to verbal memory that had strong correlations with core language measures of gold-standard clinical language assessments, providing support that our battery can objectively measure language ability. Additionally, the Lingo-derived factors showed high test-retest reliability. Analysis of Lingo factor scores offered insights into the interrelationship that language ability has with genetics, SES, and mental health. While many of these analyses cannot infer causality (i.e., low language abilities early in life lead to poor SES in adulthood or vice versa), these findings speak to the importance of considering language when studying SES and mental health. Taken together, these validations suggest that our battery is a robust platform for measuring individual differences in language/cognitive skills that are relevant to behavioral and social sciences.

The scalability of Lingo makes it a prime candidate for prospective and genetic study designs. Our analyses showing significant associations between PGS and our factor scores suggest there is a biological basis to the variation captured by Lingo, making this assessment amenable to a large-scale GWAS in the future. Finally, the ability to deploy this battery in large population samples, allowing for rapid deep phenotyping, is ideal for genetic analysis like GWAS. Follow up studies, like GWAS or rare variant association tests on Lingo scores could clarify the role genes may play in specific speech and language domains. Based on previous large-scale GWASs on reading phenotypes and the heritability of multiple language domains [4, 7], we expect that the different Lingo scores would be genetically distinct.

The recorded audio that Lingo provides is a significant benefit that distinguishes it from traditional language and cognitive assessments. Lingo's recorded audio and millisecond-scale timing allows for unparalleled feature extraction, allowing researchers to address a variety of research questions with data collected from one session. The richness of the raw recordings allows for a degree of "future proofing" as new analytic techniques are developed to derive features with greater signal-to-noise and interpretability properties. Beyond the 14 features we examined in the factor analysis, Lingo also offers the ability to assess acoustic speech qualities. Changes in speech acoustics have been linked to a variety of mental health and mood outcomes, like depression and mania in bipolar disorder [14]. It can be extremely difficult to diagnose or treat manic and depressed individuals, but tools like Lingo offer a novel avenue to collect this data without the need for them to come into a clinical setting [14]. Additionally, the nature of the COWAT and picture narration tasks would allow for semantic or sentiment analysis by looking at the specific words participants use. Recent work has shown significant differences in the semantic space of words used by individuals with schizophrenia compared to controls [20]. Ultimately, digital phenotyping tools like Lingo could reduce mental healthcare treatment barriers and allow for improved diagnosis, monitoring, and treatment.

The relatively short time to complete and inexpensive implementation of the Lingo screener makes it a promising tool for longitudinal studies. Participants can complete the Lingo battery multiple times (as long as they have access to a web browser and device with a microphone), which would allow for longitudinal tracking of speech and cognitive phenotypes. This could be particularly useful in speech biomarker studies of neurodegenerative diseases like Alzheimer's, or on the opposite end of the age spectrum to capture developmental trajectories of language acquisition and use in children.

3.1 Limitations

While we believe the Lingo battery is a significant stride forward in scalable deep phenotyping of language and cognitive abilities, there are a number of limitations with the current implementation. First and foremost, we do not cover all domains of language. For example, written language (including grammar and spelling) are not probed in Lingo - limiting its utility. Still, it may be possible to compute proxies for written grammar from the spoken grammar in the picture narration task. Additionally, while we found machine transcription to be highly accurate, human involvement is still needed for the highest levels of accuracy. Machine transcription is also biased by confounds that may be of interest to language researchers, like accent or dialect, which likely decrease transcription accuracy. Disrupted speech and vocal acoustics seen in psychiatric, neurological, and neurodevelopmental conditions could also decrease automated transcription accuracy and can lead to inaccurate results with clinical implications if they are not corrected or flagged [21]. Consequently, depending on the application, machine mistranscription may be either a feature or a bug.

4 Methods and materials

4.1 Implementation of Lingo

This language battery is a custom built, self-contained, web server. The front end (the website the participant sees and interacts with) was written with Angular (https://angular.io/), the back end was written in NodeJS (https://nodejs.org/en/), and the server is deployed on Amazon Web Services (AWS, https://aws.amazon.com/). This design allows for straightforward customization of the tasks included in the battery, where one can add or remove tasks easily prior to deployment on AWS. Once a participant navigates to the corresponding URL, their device compatibility is tested by ensuring they can listen to and record audio. Following device testing the participant begins the battery, which takes approximately 25-35 minutes for the full battery (depending on time to complete the matrix reasoning and listening comprehension tasks). We record raw audio, timing, and item selections across the battery. The raw audio for each trial undergoes text-to-speech transcription, which was checked manually through random sampling for transcription accuracy. Audio recordings of responses were initially transcribed and timestamped by OpenAI's Whisper Large-V2 text-to-speech model []. To ensure the machine transcription process was reliable, we completed manual review of >100 randomly selected audio files and found the Whisper transcriptions to be highly accurate ($\tilde{9}5\%$).

4.2 Verbal fluency task: word naming by letter

To assess verbal fluency, we prompted participants to name as many words as they could think of that start with a provided prompt letter in 30 seconds. During this 30 second interval, audio was recorded from the participant's device microphone. The prompt letters were A, C, F, L, and S. No further instruction (e.g., exclusion of proper nouns) was given. Similar word naming tasks have long been used in neurology and psychiatry to assess verbal fluency (see the FAS or COWAT [22, 23]).

4.3 Phonological memory task: sentence repetition

Phonological memory was assessed with a sentence repetition task. Prompt sentences were provided via playback of a recording of a female narrator. Prompt sentences were structurally identical to some of those on the sentence repetition task of the CELF-5 [15]. We then augmented this set with several additional sentences generated by us to raise the difficulty ceiling. In total, 12 sentences are presented in the task, in order of increasing complexity. Three seconds after the auditory prompt, the participant's response was recorded through their device microphone.

Typical scoring of in-person administration of sentence repetition tasks provides little numerical range (i.e., scored 0-3), making it challenging to assess accuracy on easy or difficult prompts where the majority of individuals will be at one extreme score. In our implementation, we provide a continuous score that is based on the bigram accuracy, i.e., the proportion of correct bigrams (consecutive word pairs) included in the participant's response. Performance in the sentence repetition task is dependent on phonological memory, yet includes other aspects of language as well (i.e., it is a combined receptive/expressive language task). Prompt sentences can be seen in Supplementary Figure S1 (with punctuation and capitalization removed).

4.4 Reading task: rapid automatized naming

Reading fluency was tested using the well-validated Rapid Automatized Naming (RAN) task. Briefly, participants are shown a grid of numbers and asked to read them aloud as quickly as they can. Our implementation includes 4 total trials, with number grid lengths of either 25 (5x5) or 36 (6x6) integer numbers. During this task, audio from the participant reading off the grid of numbers was recorded, allowing us to calculate the time spent reading and correctness through each number grid. Actual number grids used in this task can be seen in Supplementary Figure S2

4.5 Timing task: sync voice

This task aimed to evaluate participants' abilities in vocal timekeeping and rhythm. Participants were instructed to mimic a series of "La's" pronounced at fixed intervals (e.g., every 1 second) and continue the pattern for 10 seconds. The task comprised separate trials with intervals set at 0.5, 0.75, 1, and 1.25 seconds. To analyze participant performance, we examined the raw audio recordings to pinpoint instances of vocalization during the task. From these recordings, we extracted timestamps, allowing us to calculate the time intervals between each uttered "La". Subsequently, we computed several features related to consistency, error, and overall accuracy based on these timestamps. Consistency-related features were determined by fitting linear models to each participant's response data, predicting actual prompt period times based on the intervals between their "La" utterances. Participant-level R-squared values were then extracted from these models, indicating the degree to which participants' performances across trials accurately predicted the prompt period times.

4.6 Expressive language task: picture narration

During this task, participants were prompted with pictures and asked to describe what they saw for 30 seconds. No further instruction was given. During this 30 second interval, audio was recorded from the participant's device microphone. Participants completed 4 trials, and were shown the pictures in Supplementary Figure S3

4.7 Spatial receptive language task: following directions

In order to quantify the participants' ability to interpret and follow verbal directions we implemented a task similar to the "Following Directions" task found in the CELF assessment [15]. Participants were given 8 prompts of increasing difficulty, where they heard a narrator give directions to select a specific shape in a grid of shapes. We record both the time to select an item and whether that item was correct for each trial. For example, prior to seeing the prompt in Supplementary Figure S4, participants might hear "select the blue triangle underneath the red square".

4.8 Nonverbal reasoning task: matrix reasoning

Matrix reasoning is a visual-spatial problem solving task often used as a measure of nonverbal intelligence. The prompt assets for the matrix reasoning task used here were used with permission from the well-validated UK Biobank version [24], which was presented in a comparable fashion. In brief, the matrix reasoning problems involve seeing an incomplete pattern, then selecting the option that would correctly complete the pattern from six choices. This task had 14 trials of increasing difficulty. We record both the time to select an item and whether that item was correct for each trial. An example prompt is shown in Supplementary Figure S5

4.9 Sample description: SPARK

Through a research match for the SPARK for Autism study [25, 26] we recruited 1,336 adults (mean age = 41.4 years old, SD = 10.3) with rich phenotypic data. Sample demographics can be seen in Table 1 and Supplementary Figure S6. In order to establish test-retest reliability, 14 participants retook the battery two weeks after the initial test. Individuals who reported being blind or deaf were excluded due possible issues with tasks involving visual or audio stimuli (like matrix reasoning or the spatial receptive language tasks). Approximately 1/3 of this sample had been diagnosed with autism (31.2%), the other 68.8% were parents of children with autism. Many of these individuals and/or their children had extensive genetic, cognitive, and behavioral phenotype data which were used in downstream analysis to better understand the relationship relevant phenotypes and our Lingo performance results. We utilized SPARK V10 phenotype data for downstream analysis.

4.10 Sample description: in-person

We recruited 8 children from a local neurodevelopmental condition registry (devGenes, 1) who completed a gold-standard language assessment, the CELF-5, as well as our battery. We then compared CELF-5 and Lingo scores to evaluate the suitability of the included tasks. The CELF-5 was administered by trained research assistants. Due to limitations in the valid age range for the CELF-5 (valid in ages 5-21), this sample was substantially younger than the SPARK sample. All participants were younger than 21 years old.

4.11 Imputation

There was sporadically missing data in the matrix reasoning and following directions tasks due to server issues in part of the data collection phase for the SPARK sample (we believe excessive server load led to failure to save user data; no such data loss was experienced in a second wave of recruitment). Since the other five tasks recorded in full, and many participants had the majority of their data recorded, we decided to impute the missing data to improve statistical power in downstream analyses. The imputation procedure consisted of subsetting our data to individuals missing less than 33.3% of features. We then imputed the missing values using the "missRanger" R package [27], a random forest and predictive mean matching based method. We used the following parameters: "num.trees = 500", "pmm.k = 10" and a seed value for reproducibility. Imputation converged after 4 iterations, and provided a final dataset of 1,344 samples with 14 features each.

4.12 Factor analysis

Prior to determining the factor structure with exploratory factor analysis of the Lingo screener, we residualized for the effects of age and sex for each input variable. Then, we fit the initial factor model using 604 samples who had complete Lingo data (i.e., did not suffer any data loss and were not repeats). To determine the optimal factor number, we used the "fa.parallel" function from the "psych" package in R [27]. The "fa.parallel" function uses parallel analysis to identify the number of factors in a dataset by comparing eigenvalues of the observed data to eigenvalues of a random dataset with the same dimensionality. Using this data-driven method, we identified five factors in our data. We then fit our five factor model with the "fa" function from the "psych" package. After computing initial factor scores, we removed 36 outliers (i.e., samples whose absolute value of a factor score was greater than 3 * median absolute deviation + median value of that factor), and refit the five factor model. Once factors were derived, we predicted factor scores for all of the held out data (i.e., the samples with imputed data or retest data). In total, we computed Lingo factor scores for 1,336 unique SPARK participants (14 of these samples had data from 2 Lingo sessions for test-retest reliability validation) and 8 participants from our local cohort.

4.13 Genetic data and polygenic score analysis

Much of this methodology has been previously described in our prior work [28]. We describe a few small updates to our methodology and ensured we followed current best practices with the PGS calculation tool used here.

4.13.1 Genotype Quality Control and Imputation

For the SPARK reseach match cohort, many participants had genetic data available. We used the genotype arrays from SPARK integrated whole-exome sequencing (iWES1) 2022 release and SPARK whole-genome sequencing (WGS) releases 2, 3, and 4. iWES1 (n = 69,592) was quality controlled on release, including removing samples due to heterozygosity or high missingness, so no quality control was performed by us before imputation. iWES1 provided genetic ancestry assignments based on 1000 Genomes populations (26). WGS release 2 (n = 2365), release 3 (n = 2871), and release 4 (n = 3684) were not quality controlled on release, so we performed quality control using PLINK (27) before imputation. First, we removed participants from WGS if they were in iWES1. Second, we removed variants with missingness > 0.1 and participants with missingness > 0.2. Third, we merged the 3 releases and removed any participant whose heterozygosity (F statistic) was not within 3 standard deviations of the mean. We used the TopMed reference panel to identify strand flips [29]. The final sample size for WGS 2 to 4 was n = 8152. iWES1 and WGS 2 to 4 were then imputed to TopMed [29] using the Michigan Imputation Server [30] with phasing and quality control steps included and to output variants with imputation quality r2 ¿ 0.3. After imputation, variants were filtered to single nucleotide polymorphisms with imputation quality r2 ¿ 0.8 and a minor allele frequency > 0.1% in our sample using bcftools [31]. They were lifted over from hg38 to hg19 using the VCF-liftover tool (https://github.com/hmgu-itg/VCF-liftover) and normalized to the hg19 reference genome. Finally, files were merged,

subset to HapMap3 Plus variants [32], and variants with 0% missingness were retained (final number of variants = 1,147,121).

4.13.2 Genetic Ancestry

Genetic principal components (PCs) were calculated using bigsnpr [33], specifically following the authors' recommendations [34] and tutorial: https://privefl.github.io/bigsnpr/articles/bedpca.html. In summary, we 1) used the "snpplinkKINGQC" function to identify and remove related participants at KING threshold of 2-3.5, 2) performed PC analysis using the bedautoSVD with unrelated participants, 3) detected and removed PC outliers, 4) recalculated PCs, and 5) projected PCs onto the entire cohort using the bedprojectSelfPCA function. We used 40 PCs and performed k-means clustering with K = 5 (for the 5 populations of 1000 Genomes [35] and used the genetic ancestry labels from iWES1 to assign labels to the genetic population clusters. Due to the lack of diversity in many GWASs, PGS based on these results tend to perform much better in individuals with predominantly European ancestry. To ensure our PGS were valid and accurate, for the PGS analysis we subset to individuals who clustered with the Europeans samples (final N = 811 with QC passing genetic data, were European, and had Lingo data).

4.13.3 Polygenic Scores

PGSs were calculated using the "LDpred2" [36] and "bigsnpr" tools [33] in R. Because SPARK is family based, an external linkage disequilibrium reference based on 362,320 participants in the UK Biobank (provided by the authors of LDpred2) was used to calculate infinitesimal beta weights. PGSs were calculated from the following GWASs conducted by the Psychiatric Genetics Consortium: alcohol abuse [37], Alzheimer's disease [38], attention-deficit/hyperactivity disorder (ADHD) [39], anorexia nervosa [40], autism [41], bipolar disorder [42], major depression [43], post traumatic stress disorder (PTSD) [44], and schizophrenia [45]. As well as the following GWASs conducted by the the Social Science Genetic Association Consortium: cognitive performance [46], educational attainment [47], and risk taking [48]. To correct for population stratification we residualized out the main effects of the first 20 genetic PCs from each PGS and z-scaled the scores. Finally, we tested for PGS associations with Lingo factor scores using Pearson's correlations, as both the PGS and Lingo scores were normally distributed. P-values were adjusted for multiple testing burden by using the False Discovery Rate (FDR) method with the "p.adjust" function in R.

4.14 Statistical analysis

All statistical analysis was conducted in R version 4.3.0 [49].

Acknowledgments

We are grateful to all of the participants and families in SPARK, the SPARK research match team and Tempus, the SPARK clinical sites, and SPARK staff. We appreciate obtaining access to genetic and phenotypic data for SPARK data on SFARI Base. We are very appreciative of time and effort of the research assistants who helped quality control the data. Some of the study data were collected and managed using REDCap electronic data capture tools hosted at the Univerity of Iowa.

Data availability

The SPARK genetic and phenotypic data can be obtained at SFARI Base: `https://base.sfari.org`
The SPARK Research Match data will be available to qualified, approved researchers through SFARI Base upon publication of this article.

Funding

This work was supported by the National Institutes of Health (DC014489 to JJM) as well as grants from the Simons Foundation (SFARI 516716 to JJM) and the Clinical and Translational Science Award (KL2TR001877 to JJM). This work was also supported by the Roy J. Carver Charitable Trust through a professorship to JJM. This work was supported by the University of Iowa Hawkeye Intellectual and Developmental Disabilities Research Center (Hawk-IDDRC) through the Eunice Kennedy Shriver National Institute of Child Health and Human Development (P50HD103556).

Conflicts of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author contributions

Lingo was originally conceived by JJM with input from JBT and the initial version developed by MS. Further conceptual and software development was contributed by TK, CM, BT, and LGC. The study was designed by LGC and JJM. The SPARK Research Match study was carried out by TK and LGC with guidance by JJM. The data was processed by LGC and ME. The analyses were performed by LGC and JJM. The manuscript was drafted by LGC and JJM and was approved by all authors.

Figures and Tables

sample	N	Mean age (std. dev.)	Females	Males	ASD
SPARK	1336	41.43 (10.3)	1085 (81.2%)	251 (18.7%)	423 (31.7%), 132 male
SPARK retest	14	43.3 (6.5)	11 (78.6%)	3 (21.4%)	3 (21.4%), 1 male
Local cohort (devGenes)	8	13.97 (1.8)	2 (25%)	6 (75%)	2 (25%), 2 male

Table 1. Demographic information of samples collected.

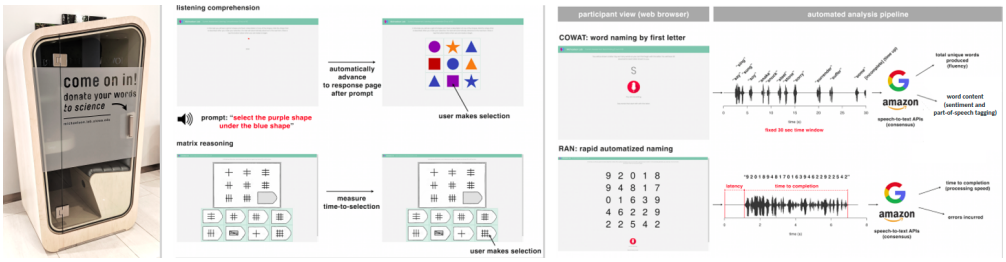

Figure 1. Schematic overview of Lingo, our online language battery. Participants were routed to the Lingo website where they completed the Lingo battery using personal computers and laptops. Raw audio, selections, and timing are recorded during Lingo. Some screenshots of the assessment are shown here as well as examples of how the data can be analyzed

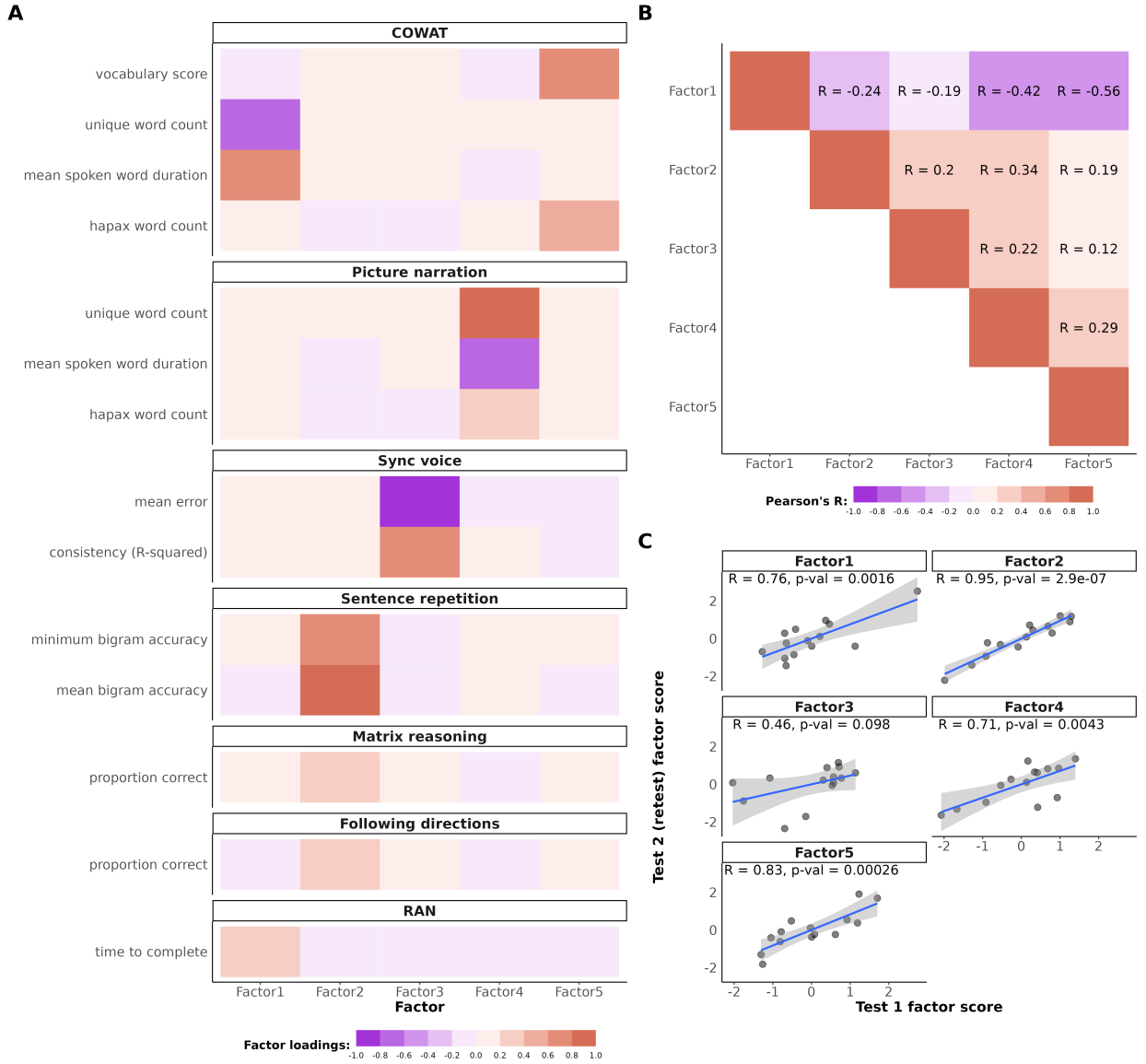

Figure 2. Factor structure of Lingo and test-retest reliability. 2.A. Factor loadings of the 14 Lingo performance features we analyzed. We found a 5 factor model fit our data best. **2.B.** Correlation heatmap between the Lingo factor scores for all participants. **2.C.** Test-retest reliability analysis of the 5 Lingo factors for the subset of participants who completed the assesment twice (at least 2 weeks apart).

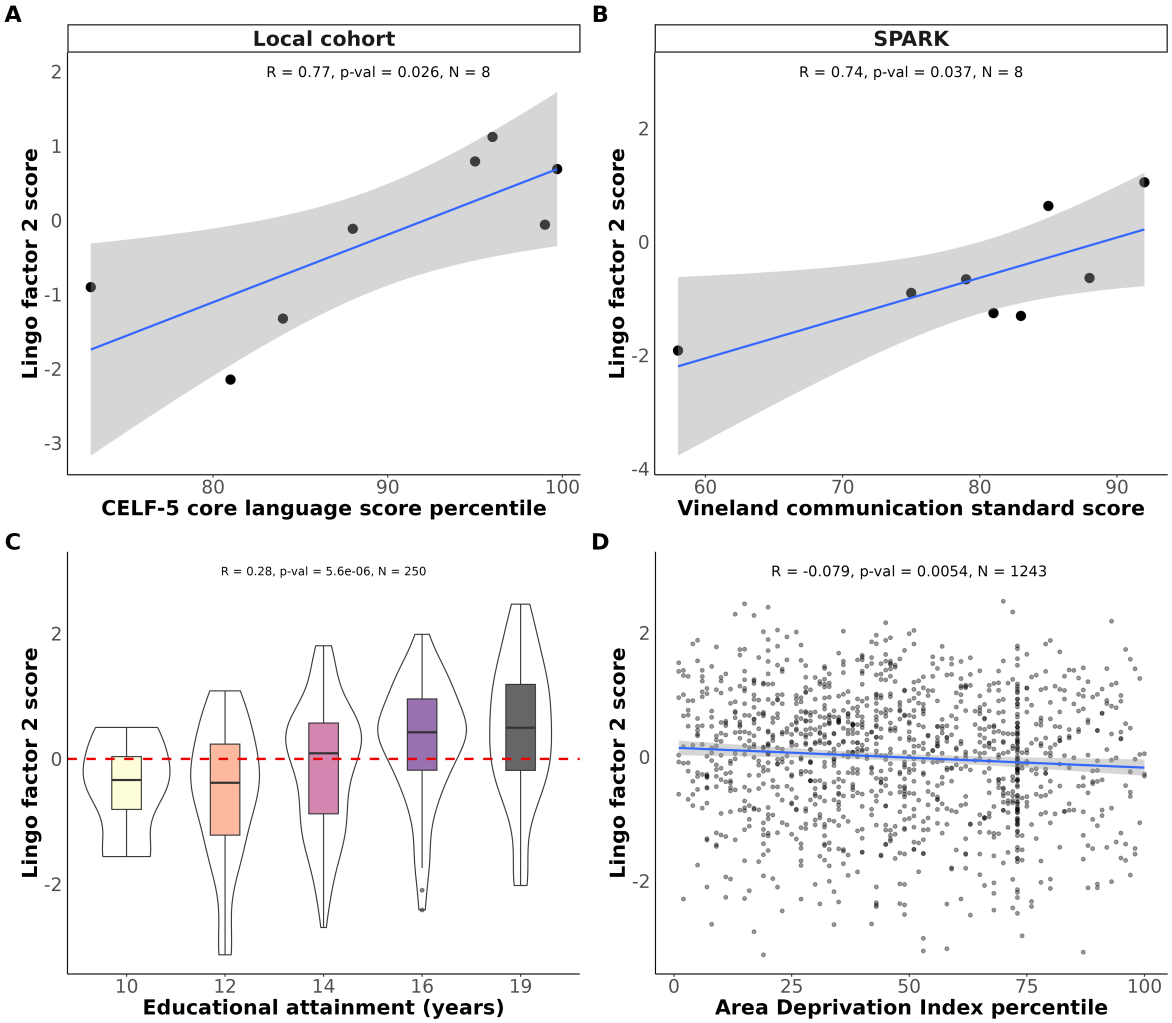

Figure 3. Validation of Lingo factors with gold-standard assessments and SES. 3.A. Concurrent validity analysis of the Lingo factor 2 scores with the CELF-5 scores in a small local cohort who completed both assessments. **3.B.** Concurrent validity analysis of the Lingo factor 2 scores with the Vineland Adaptive Behavior Scales-3 language score in a small subset of SPARK participants who had data for both assessments available. **3.C.** Correlation between educational attainment in years (i.e., 10 = did not complete high school, 19 = graduate or professional degree) and Lingo factor 2 scores. As many young adults are still in school, we subset the sample to individuals > 30 years old to limit the effects of age in our analysis. **3.D.** Correlation between the Area Deprivation Index and Lingo factor 2 scores.

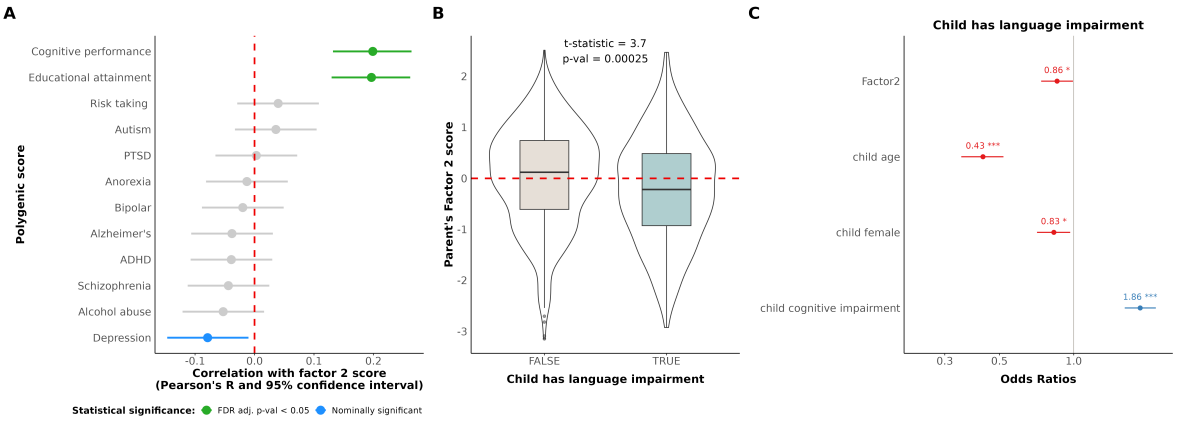

Figure 4. Genetic analysis of Lingo factors. 4.A. Brain related polygenic score (PGS) correlations with Lingo factor 2 scores for European SPARK participants with available genetic data (N = 811). Due to the multiple testing burden we used FDR p-value correction. Plotted are the Pearson's R and 95% confidence interval estimates for correlations between PGS and Lingo factor 2 scores. **4.B.** Parental score on Lingo factor 2 was associated with their child's language impairment. **4.C.** Parental score on Lingo factor 2 are an independent predictor of their child's language impairment after accounting for their child's cognitive impairment, age, and sex. Odds' ratios and 95% confidence interval computed from a multivariate logisitc regression model predicting the child's language impairment are shown. "***" indicates p-value < 0.001, "**" indicates p-value < 0.01, and "*" indicates p-value < 0.05.

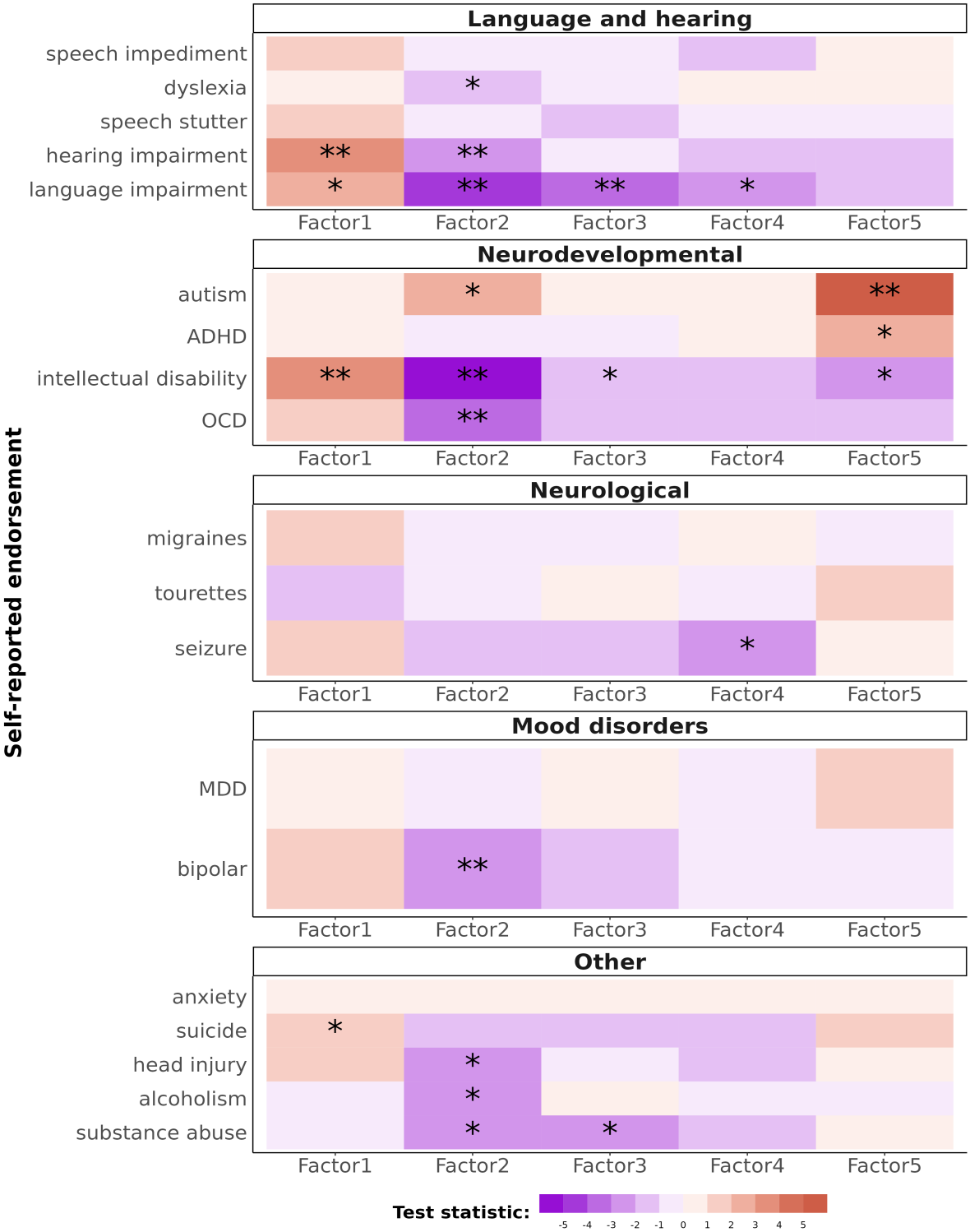

Figure 5. Lingo factors associations with psychiatric and neurodevelopmental conditions. Plotted are the test statistics from logistic regression models predicting self-reported diagnosis from the Lingo factor scores after accounting for age and sex (N = 845). Due to the multiple testing burden we used FDR p-value correction. "**" indicates FDR adjusted p-values are less than 0.05. "*" indicates nominal statistical significance, unadjusted p-vale < 0.05 but the FDR adjusted p-value > 0.05. Negative values indicate scoring high on the Lingo factor is associated with a lower likelihood of having that diagnosis. Positive values indicate scoring high on the Lingo factor is associated with a higher likelihood of having that diagnosis. Values near zero indicate no significant association.

Supplementary information

Supplementary tables and figures

Task	Feature	Description
Verbal fluency task: **Controlled Oral Word Association Test (COWAT)**	Vocabulary score	A score designed to measure the average word frequency of words said by the participant. Word frequencies were computed in sample. This score is the -log10(mean word frequency). High scores indicate the participant is consistently using words most other people are not, and may indicate a high vocabulary.
	Unique word count	Number of unique words in a response. (Estimate for verbal fluency)
	Mean time to say word	Average time taken to say a word, a proxy of word length and verbal fluency
	Hapax word count	A hapax is a word that was said only once in our sample. This score represents the total number of hapaxes said by the participant.
Expressive language task: **Picture narration**	Unique word count	Number of unique words in narratives by a participant
	Mean time to say word	Average time taken to say a word, a proxy of word length and verbal fluency
	Hapax word count	A hapax is a word that was said only once in our sample. This score represents the totoal number of hapaxes said by the participant.
Rhythm/timekeeping task: **Sync voice**	Consistency (R-squared)	Participant-level R-squared values from linear models to predict accuracy
	Mean error	Participant-level mean error to predict accuracy
Phonological memory task: **Sentence repetition**	Minimum bigram accuracy	Minimum accuracy of bigrams (consecutive word pairs) in all trials
	Mean bigram accuracy	Average of bigram accuracy.
Nonverbal reasoning task: **Matrix reasoning**	Proportion correct	Proportion of correct trials
Spatial receptive language task: **Following directions**	Proportion correct	Proportion of correct trials
Reading task: **Rapid automatized naming (RAN)**	Time to complete	Average time it took participants to complete the RAN trials (in seconds)

Table 2. Description of Lingo features.

prompt number	prompt
1	didnt the girl paint the picture
2	the hamster was not left in the cage
3	the student did not see the teacher who taught first grade last year
4	if the driver had delivered some donuts they would have been eaten by now
5	the woman who helps serve the hot lunches at school is the new nurse
6	the girl who didnt participate in practice wasnt allowed to start for the team until a month later
7	the art teacher prepared printed organized and delivered the new sketchbooks
8	the janitor from down the hall promised to feed the python during our winter vacation
9	it is the teachers practice to always assign homework two days before it is due and no more than three days following the previous assignment
10	the cafeteria doors around which the seventh graders sometimes stand until the lunch bell rings are always closed but never locked
11	by secretly helping herself to his extra pencils she was ensuring neither would have one available when the big test he had been worrying about was finally given
12	having been tardy more days than not the brothers both knew it was only a matter of time before each would have to be responsible for himself

Figure S1. Prompt sentences used in the sentence repetition task (with punctuation and capitalization removed).

Prompt: 1

1	2	5	2	3
5	5	3	2	3
1	1	1	3	5
2	5	1	3	5
1	2	3	2	3

Prompt: 2

9	2	0	1	8
9	4	8	1	7
0	1	6	3	9
4	6	2	2	9
2	2	5	4	2

Prompt: 3

5	4	2	6	3	1
8	2	6	7	5	2
5	3	2	8	9	8
3	7	1	6	7	8
4	3	6	7	1	3
6	7	5	7	7	5

Prompt: 4

4	7	1	3	1	5
2	7	6	9	1	8
2	3	7	8	1	1
3	2	7	7	4	4
2	5	2	9	6	1
4	1	7	2	1	7

Figure S2. **Prompt number grids used in the RAN reading task.**

Figure S3. **Please contact the authors for a copy of the prompt images. Prompt images used in the picture narration task will be included here in the final peer reviewed version of the manuscript.**

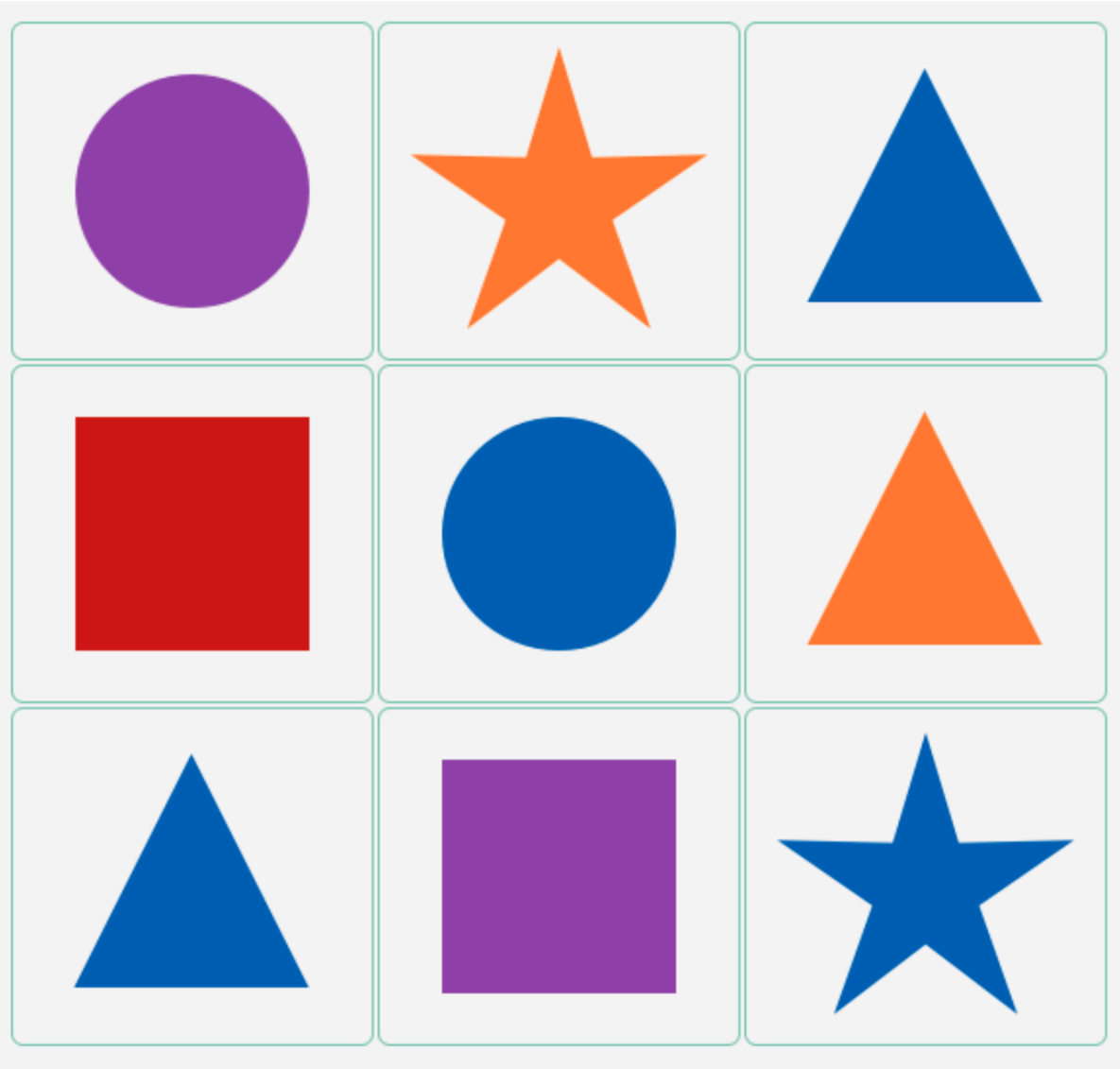

Figure S4. Prompt 1 image for the following directions. Participants were asked to click on the shape that matched the verbal description they had heard.

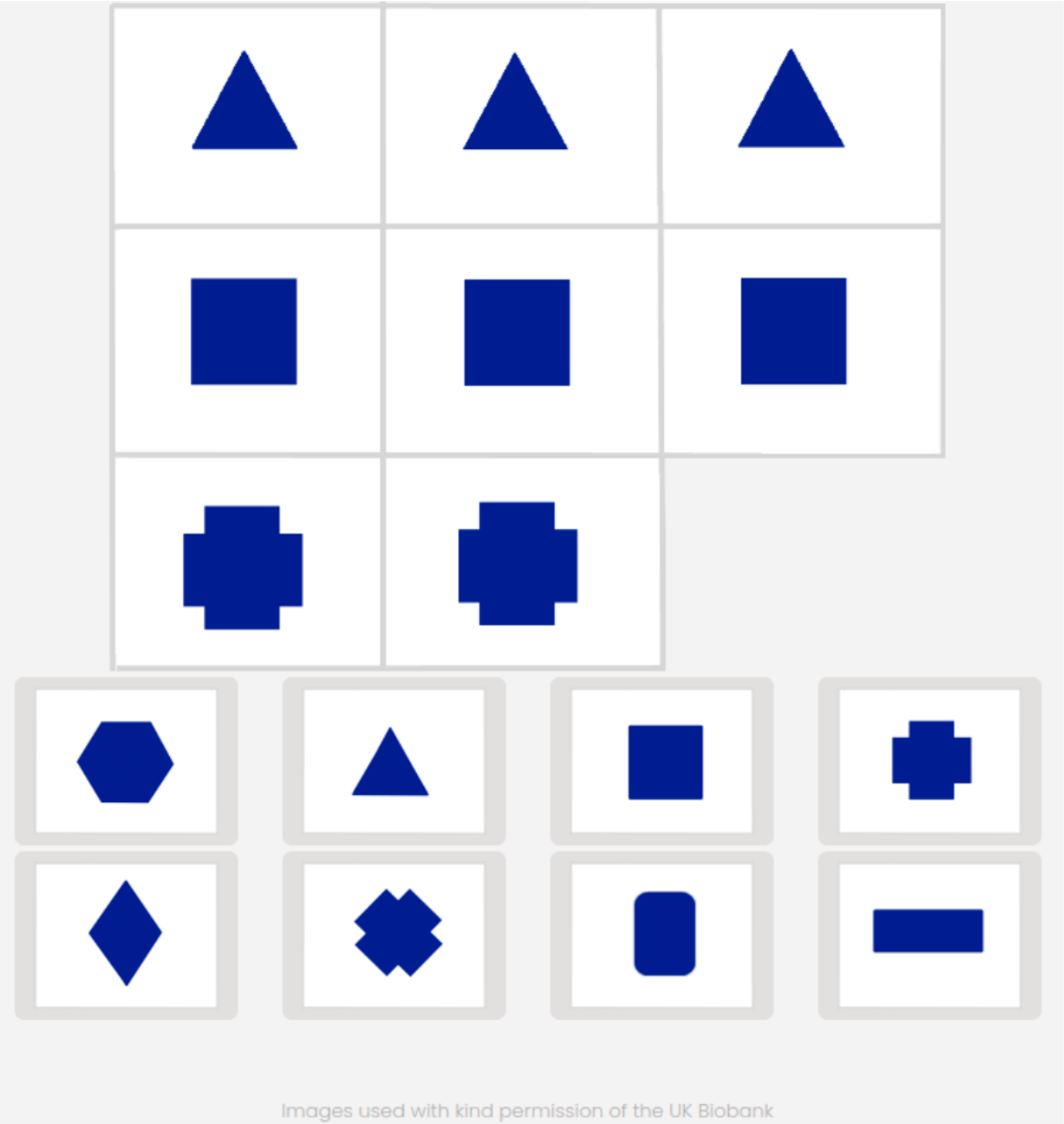

Figure S5. Prompt 1 image for the matrix reasoning task, used with permission from same task in the UK Biobank. Participants were asked to click on the pattern that best fit the matrix shown above.

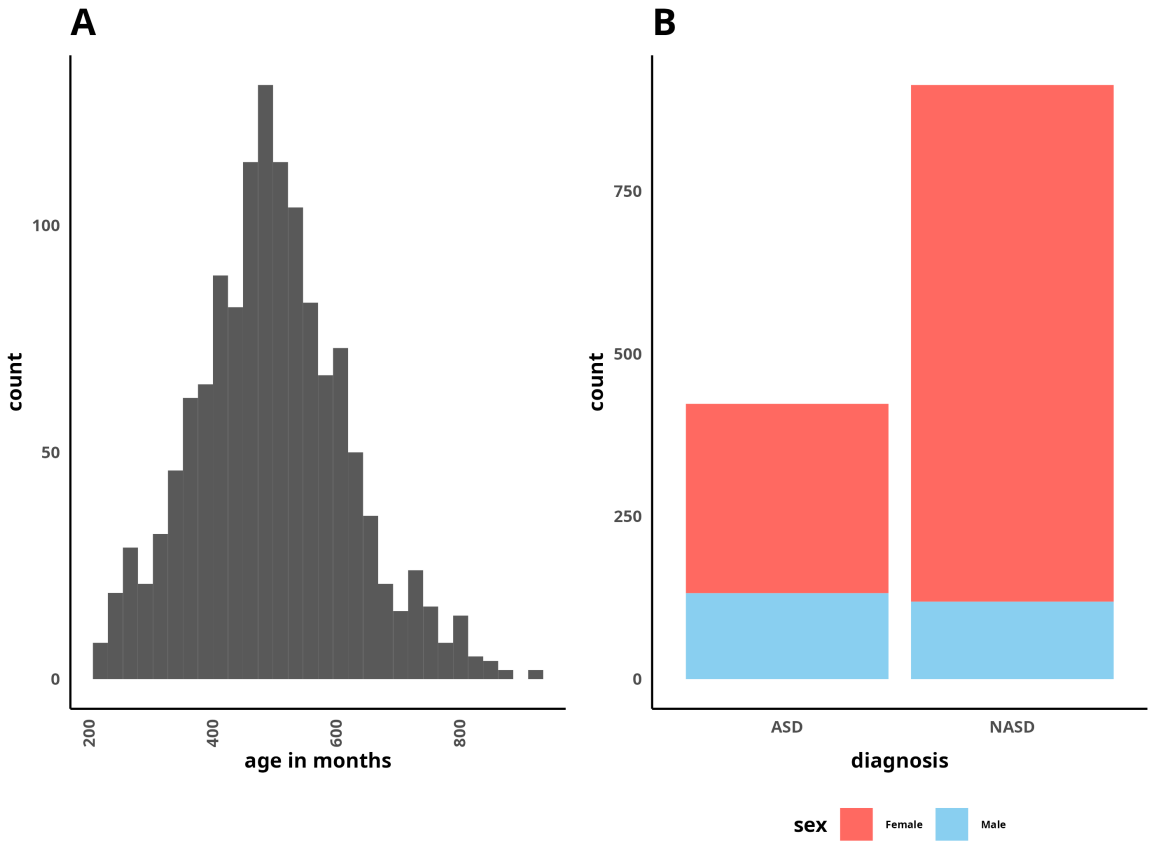

Figure S6. Demographic distributions of the SPARK Lingo cohort.

References

1. E. Fedorenko, "The role of domain-general cognitive control in language comprehension," *Frontiers in Psychology*, vol. 5, 4 2014.
2. Z. Deldar, C. Gevers-Montoro, A. Khatibi, and L. Ghazi-Saidi, "The interaction between language and working memory: a systematic review of fmri studies in the past two decades.," *AIMS neuroscience*, vol. 8, pp. 1–32, 2021.
3. A. C. Yu and G. Zellou, "Individual differences in language processing: Phonology," *Annual Review of Linguistics*, vol. 5, pp. 131–150, 1 2019.
4. D. V. M. Bishop, G. Laws, C. Adams, and C. F. Norbury, "High heritability of speech and language impairments in 6-year-old twins demonstrated using parent and teacher report," *Behavior Genetics*, vol. 36, pp. 173–184, 3 2006.
5. J. B. Tomblin and P. R. Buckwalter, "Heritability of poor language achievement among twins," *Journal of Speech, Language, and Hearing Research*, vol. 41, pp. 188–199, 2 1998.
6. S. A. Graham and S. E. Fisher, "Understanding language from a genomic perspective," *Annual Review of Genetics*, vol. 49, pp. 131–160, 11 2015.
7. E. Eising, N. Mirza-Schreiber, E. L. de Zeeuw, C. A. Wang, D. T. Truong, A. G. Allegrini, C. Y. Shapland, G. Zhu, K. G. Wigg, M. L. Gerritse, B. Molz, G. Alagöz, A. Gialluisi, F. Abbondanza, K. Rimfeld, M. van Donkelaar, Z. Liao, P. R. Jansen, T. F. M. Andlauer, T. C. Bates, M. Bernard, K. Blokland, M. Bonte, A. D. Børglum, T. Bourgeron, D. Brandeis, F. Ceroni, V. Csépe, P. S. Dale, P. F. de Jong, J. C. DeFries, J.-F. Démonet, D. Demontis, Y. Feng, S. D. Gordon, S. L. Guger, M. E. Hayiou-Thomas, J. A. Hernández-Cabrera, J.-J. Hottenga, C. Hulme, J. Kere, E. N. Kerr, T. Koomar, K. Landerl, G. T. Leonard, M. W. Lovett, H. Lyytinen, N. G. Martin, A. Martinelli, U. Maurer, J. J. Michaelson, K. Moll, A. P. Monaco, A. T. Morgan, M. M. Nöthen, Z. Pausova, C. E. Pennell, B. F. Pennington, K. M. Price, V. M. Rajagopal, F. Ramus, L. Richer, N. H. Simpson, S. D. Smith, M. J. Snowling, J. Stein, L. J. Strug, J. B. Talcott, H. Tiemeier, M. P. van der Schroeff, E. Verhoef, K. E. Watkins, M. Wilkinson, M. J. Wright, C. L. Barr, D. I. Boomsma, M. Carreiras, M.-C. J. Franken, J. R. Gruen, M. Luciano, B. Müller-Myhsok, D. F. Newbury, R. K. Olson, S. Paracchini, T. Paus, R. Plomin, S. Reilly, G. Schulte-Körne, J. B. Tomblin, E. van Bergen, A. J. O. Whitehouse, E. G. Willcutt, B. S. Pourcain, C. Francks, and S. E. Fisher, "Genome-wide analyses of individual differences in quantitatively assessed reading- and language-related skills in up to 34,000 people," *Proceedings of the National Academy of Sciences*, vol. 119, 8 2022.
8. M. Malanchini, K. Rimfeld, A. Gidziela, R. Cheesman, A. G. Allegrini, N. Shakeshaft, K. Schofield, A. Packer, R. Ogden, A. McMillan, S. J. Ritchie, P. S. Dale, T. C. Eley, S. von Stumm, and R. Plomin, "Pathfinder: a gamified measure to integrate general cognitive ability into the biological, medical, and behavioural sciences," *Molecular Psychiatry*, vol. 26, pp. 7823–7837, 12 2021.
9. C. J. Johnson, J. H. Beitchman, A. Young, M. Escobar, L. Atkinson, B. Wilson, E. B. Brownlie, L. Douglas, N. Taback, I. Lam, and M. Wang, "Fourteen-year follow-up of children with and without speech/language impairments," *Journal of Speech, Language, and Hearing Research*, vol. 42, pp. 744–760, 6 1999.
10. E. J. Tan, G. W. Yelland, and S. L. Rossell, "Characterising receptive language processing in schizophrenia using word and sentence tasks," *Cognitive Neuropsychiatry*, vol. 21, pp. 14–31, 1 2016.
11. S. E. Morgan, K. Diederen, P. E. Vértes, S. H. Y. Ip, B. Wang, B. Thompson, A. Demjaha, A. D. Micheli, D. Oliver, M. Liakata, P. Fusar-Poli, T. J. Spencer, and P. McGuire, "Natural language processing markers in first episode psychosis and people at clinical high-risk," *Translational Psychiatry*, vol. 11, p. 630, 12 2021.
12. M. Gumus, D. D. DeSouza, M. Xu, C. Fidalgo, W. Simpson, and J. Robin, "Evaluating the utility of daily speech assessments for monitoring depression symptoms.," *Digital health*, vol. 9, p. 20552076231180523, 2023.
13. D. K. Anderson, C. Lord, S. Risi, P. S. DiLavore, C. Shulman, A. Thurm, K. Welch, and A. Pickles, "Patterns of growth in verbal abilities among children with autism spectrum disorder.," *Journal of consulting and clinical psychology*, vol. 75, pp. 594–604, 8 2007.
14. D. M. Low, K. H. Bentley, and S. S. Ghosh, "Automated assessment of psychiatric disorders using speech: A systematic review," *Laryngoscope Investigative Otolaryngology*, vol. 5, pp. 96–116, 2 2020.
15. E. H. Wiig, W. A. Secord, and E. Semel, *Clinical evaluation of language fundamentals: CELF-5*. Pearson, 2013.
16. S. S. Sparrow, D. V. Ciccheti, and C. A. Saulnier, *Vineland Adaptive Behavior Scales, third edition*. Pearson, 2013.
17. D. Tomasi and N. D. Volkow, "Associations of family income with cognition and brain structure in usa children: prevention implications," *Molecular Psychiatry*, vol. 26, pp. 6619–6629, 11 2021.
18. K. G. Noble, M. F. Norman, and M. J. Farah, "Neurocognitive correlates of socioeconomic status in kindergarten children," *Developmental Science*, vol. 8, pp. 74–87, 1 2005.
19. C. Burneo-Garcés, F. Cruz-Quintana, M. Pérez-García, M. Fernández-Alcántara, A. Fasfous, and M. N. Pérez-Marfil, "Interaction between socioeconomic status and cognitive development in children aged 7, 9, and 11 years: A cross-sectional study," *Developmental Neuropsychology*, vol. 44, pp. 1–16, 1 2019.
20. M. M. Nour, D. C. McNamee, Y. Liu, and R. J. Dolan, "Trajectories through semantic spaces in schizophrenia and the relationship to ripple bursts," *Proceedings of the National Academy of Sciences*, vol. 120, 10 2023.
21. T. Soroski, T. da Cunha Vasco, S. Newton-Mason, S. Granby, C. Lewis, A. Harisinghani, M. Rizzo, C. Conati, G. Murray, G. Carenini, T. S. Field, and H. Jang, "Evaluating web-based automatic transcription for alzheimer speech data: Transcript comparison and machine learning analysis," *JMIR Aging*, vol. 5, p. e33460, 9 2022.
22. J. R. Kaufman, H. Fatima, L. H. Lacritz, and C. M. Cullum, "Utility of a short-form phonemic fluency task.," *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists*, 3 2024.
23. R. M. Ruff, R. H. Light, S. B. Parker, and H. S. Levin, "Benton controlled oral word association test: reliability and updated norms.," *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists*, vol. 11, pp. 329–38, 1996.
24. C. Fawns-Ritchie and I. J. Deary, "Reliability and validity of the uk biobank cognitive tests," *PLOS ONE*, vol. 15, p. e0231627, 4 2020.

25. P. Feliciano, A. M. Daniels, L. G. Snyder, A. Beaumont, A. Camba, A. Esler, A. G. Gulsrud, A. Mason, A. Gutierrez, A. Nicholson, A. M. Paolicelli, A. P. McKenzie, A. L. Rachubinski, A. N. Stephens, A. R. Simon, A. Stedman, A. D. Shocklee, A. Swanson, B. Finucane, B. A. Hilscher, B. Hauf, B. J. O'Roak, B. McKenna, B. E. Robertson, B. Rodriguez, B. M. Vernoia, B. V. Metre, C. Bradley, C. Cohen, C. A. Erickson, C. Harkins, C. Hayes, C. Lord, C. L. Martin, C. Ortiz, C. Ochoa-Lubinoff, C. Peura, C. E. Rice, C. R. Rosenberg, C. J. Smith, C. Thomas, C. M. Taylor, L. C. White, C. H. Walston, D. G. Amaral, D. L. Coury, D. E. Sarver, D. Istephanous, D. Li, D. C. Nugyen, E. A. Fox, E. M. Butter, E. Berry-Kravis, E. Courchesne, E. J. Fombonne, E. Hofammann, E. Lamarche, E. L. Wodka, E. T. Matthews, E. O'Connor, E. Palen, F. Miller, G. S. Dichter, G. Marzano, G. Stein, H. Hutter, H. E. Kaplan, H. Li, H. Lechniak, H. L. Schneider, H. Zaydens, I. Arriaga, J. A. Gerdts, J. F. Cubells, J. M. Cordova, J. Gunderson, J. Lillard, J. Manoharan, J. T. McCracken, J. J. Michaelson, J. Neely, J. Orobio, J. Pandey, J. Piven, J. Scherr, J. S. Sutcliffe, J. Tjernagel, J. Wallace, K. Callahan, K. Dent, K. A. Schweers, K. E. Hamer, J. K. Law, K. Lowe, K. O'Brien, K. Smith, K. G. Pawlowski, K. L. Pierce, K. Roeder, L. J. Abbeduto, L. N. Berry, L. A. Cartner, L. A. Coppola, L. Carpenter, L. Cordeiro, L. DeMarco, L. P. Grosvenor, L. Higgins, L. Y. Huang-Storms, L. Hosmer-Quint, L. M. Herbert, L. Kasparson, L. M. Prock, L. D. Pacheco, L. Raymond, L. Simon, L. V. Soorya, L. Wasserburg, M. Lazar, M. Alessandri, M. Brown, M. H. Currin, M. F. Gwynette, M. Heyman, M. N. Hale, M. Jones, M. Jordy, M. J. Morrier, M. Sahin, M. S. Siegel, M. Verdi, M. V. Parlade, M. Yinger, N. Bardett, N. Hanna, N. Harris, N. Pottschmidt, N. Russo-Ponsaran, N. Takahashi, O. Y. Ousley, A. P. Juarez, P. Manning, R. D. Annett, R. A. Bernier, R. D. Clark, R. J. Landa, R. P. Goin-Kochel, R. Remington, R. T. Schultz, S. J. Brewster, S. Booker, S. Carpenter, S. Eldred, S. Francis, S. L. Friedman, S. Horner, S. Hepburn, S. Jacob, S. Kanne, S. J. Lee, S. A. Mastel, S. Plate, S. Qiu, S. Sandhu, S. Thompson, S. White, V. J. Myers, V. Singh, W. S. Yang, Z. Warren, A. Amatya, A. J. Ace, A. S. Chatha, A. E. Lash, B. Negron, C. Rigby, C. Ridenour, C. M. Stock, D. Schmidt, I. Fisk, J. Acampado, J. L. Nestle, J. A. Nestle, K. Layman, M. E. Butler, M. Kent, M. D. Mallardi, N. Carriero, N. Lawson, N. Volfovsky, R. Edgar, R. Marini, R. Rana, S. Ganesan, S. Shah, T. Ramsey, W. Chin, W. Jensen, A. D. Krentz, A. J. Gruber, A. Sabo, A. Salomatov, C. Eng, D. Muzny, I. Astrovskaya, R. A. Gibbs, X. Han, Y. Shen, L. F. Reichardt, and W. K. Chung, "Spark: A us cohort of 50,000 families to accelerate autism research," *Neuron*, vol. 97, pp. 488–493, 2 2018.

26. K. K. Sudnawa and W. K. Chung, "Sparking new insight into autism across the lifespan," *American Journal on Intellectual and Developmental Disabilities*, vol. 129, pp. 91–95, 3 2024.

27. M. Mayer, *missRanger: Fast Imputation of Missing Values*, 2024. R package version 2.4.0, https://mayer79.github.io/missRanger/.

28. T. R. Thomas, A. J. Tener, A. M. Pearlman, K. L. Imborek, J. S. Yang, J. F. Strang, and J. J. Michaelson, "Polygenic scores clarify the relationship between mental health and gender diversity," *Biological Psychiatry Global Open Science*, vol. 4, p. 100291, 3 2024.

29. D. Taliun, D. N. Harris, M. D. Kessler, J. Carlson, Z. A. Szpiech, R. Torres, S. A. G. Taliun, A. Corvelo, S. M. Gogarten, H. M. Kang, *et al.*, "Sequencing of 53,831 diverse genomes from the nhlbi topmed program," *Nature*, vol. 590, pp. 290–299, 2021.

30. S. Das, L. Forer, S. Schönherr, C. Sidore, A. E. Locke, A. Kwong, S. I. Vrieze, E. Y. Chew, S. Levy, M. McGue, D. Schlessinger, D. Stambolian, P.-R. Loh, W. G. Iacono, A. Swaroop, L. J. Scott, F. Cucca, F. Kronenberg, M. Boehnke, G. R. Abecasis, and C. Fuchsberger, "Next-generation genotype imputation service and methods.," *Nature genetics*, vol. 48, pp. 1284–1287, 2016.

31. P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, *et al.*, "Twelve years of SAMtools and BCFtools," *Gigascience*, vol. 10, no. 2, p. giab008, 2021.

32. F. Privé, C. Albiñana, J. Arbel, B. Pasaniuc, and B. J. Vilhjálmsson, "Inferring disease architecture and predictive ability with ldpred2-auto," *The American Journal of Human Genetics*, vol. 110, pp. 2042–2055, 12 2023.

33. F. Privé, H. Aschard, A. Ziyatdinov, and M. G. B. Blum, "Efficient analysis of large-scale genome-wide data with two r packages: bigstatsr and bigsnpr," *Bioinformatics*, vol. 34, pp. 2781–2787, 2018.

34. F. Privé, K. Luu, M. G. Blum, J. J. McGrath, and B. J. Vilhjálmsson, "Efficient toolkit implementing best practices for principal component analysis of population genetic data," *Bioinformatics*, vol. 36, no. 16, pp. 4449–4457, 2020.

35. . G. P. Consortium *et al.*, "A global reference for human genetic variation," *Nature*, vol. 526, no. 7571, p. 68, 2015.

36. F. Privé, J. Arbel, and B. J. Vilhjálmsson, "Ldpred2: better, faster, stronger," *Bioinformatics*, vol. 36, pp. 5424–5431, 2020.

37. R. K. Walters, R. Polimanti, E. C. Johnson, J. N. McClintick, M. J. Adams, A. E. Adkins, F. Aliev, S.-A. Bacanu, A. Batzler, S. Bertelsen, J. M. Biernacka, T. B. Bigdeli, L.-S. Chen, T.-K. Clarke, Y.-L. Chou, F. Degenhardt, A. R. Docherty, A. C. Edwards, P. Fontanillas, J. C. Foo, L. Fox, J. Frank, I. Giegling, S. Gordon, L. M. Hack, A. M. Hartmann, S. M. Hartz, S. Heilmann-Heimbach, S. Herms, C. Hodgkinson, P. Hoffmann, J. J. Hottenga, M. A. Kennedy, M. Alanne-Kinnunen, B. Konte, J. Lahti, M. Lahti-Pulkkinen, D. Lai, L. Ligthart, A. Loukola, B. S. Maher, H. Mbarek, A. M. McIntosh, M. B. McQueen, J. L. Meyers, Y. Milaneschi, T. Palviainen, J. F. Pearson, R. E. Peterson, S. Ripatti, E. Ryu, N. L. Saccone, J. E. Salvatore, S. Sanchez-Roige, M. Schwandt, R. Sherva, F. Streit, J. Strohmaier, N. Thomas, J.-C. Wang, B. T. Webb, R. Wedow, L. Wetherill, A. G. Wills, J. D. Boardman, D. Chen, D.-S. Choi, W. E. Copeland, R. C. Culverhouse, N. Dahmen, L. Degenhardt, B. W. Domingue, S. L. Elson, M. A. Frye, W. Gäbel, C. Hayward, M. Ising, M. Keyes, F. Kiefer, J. Kramer, S. Kuperman, S. Lucae, M. T. Lynskey, W. Maier, K. Mann, S. Männistö, B. Müller-Myhsok, A. D. Murray, J. I. Nurnberger, A. Palotie, U. Preuss, K. Räikkönen, M. D. Reynolds, M. Ridinger, N. Scherbaum, M. A. Schuckit, M. Soyka, J. Treutlein, S. Witt, N. Wodarz, P. Zill, D. E. Adkins, J. M. Boden, D. I. Boomsma, L. J. Bierut, S. A. Brown, K. K. Bucholz, S. Cichon, E. J. Costello, H. de Wit, N. Diazgranados, D. M. Dick, J. G. Eriksson, L. A. Farrer, T. M. Foroud, N. A. Gillespie, A. M. Goate, D. Goldman, R. A. Grucza, D. B. Hancock, K. M. Harris, A. C. Heath, V. Hesselbrock, J. K. Hewitt, C. J. Hopfer, J. Horwood, W. Iacono, E. O. Johnson, J. A. Kaprio, V. M. Karpyak, K. S. Kendler, H. R. Kranzler, K. Krauter, P. Lichtenstein, P. A. Lind, M. McGue, J. MacKillop, P. A. F. Madden, H. H. Maes, P. Magnusson, N. G. Martin, S. E. Medland, G. W. Montgomery, E. C. Nelson, M. M. Nöthen, A. A. Palmer, N. L. Pedersen, B. W. J. H. Penninx, B. Porjesz, J. P. Rice, M. Rietschel, B. P. Riley, R. Rose, D. Rujescu, P.-H. Shen, J. Silberg, M. C. Stallings, R. E. Tarter, M. M. Vanyukov, S. Vrieze, T. L. Wall, J. B. Whitfield, H. Zhao, B. M. Neale, J. Gelernter, H. J. Edenberg, and A. Agrawal, "Transancestral gwas of alcohol dependence reveals common genetic underpinnings with psychiatric disorders," *Nature Neuroscience*, vol. 21, pp. 1656–1669, 12 2018.

38. D. P. Wightman, I. E. Jansen, J. E. Savage, A. A. Shadrin, S. Bahrami, D. Holland, A. Rongve, S. Børte, B. S. Winsvold, O. K. Drange, A. E. Martinsen, A. H. Skogholt, C. Willer, G. Bråthen, I. Bosnes, J. B. Nielsen, L. G. Fritsche, L. F. Thomas, L. M. Pedersen, M. E. Gabrielsen, M. B. Johnsen, T. W. Meisingset, W. Zhou, P. Proitsi, A. Hodges, R. Dobson, L. Velayudhan, K. Heilbron, A. Auton, M. Agee, S. Aslibekyan, E. Babalola, R. K. Bell, J. Bielenberg, K. Bryc, E. Bullis, B. Cameron, D. Coker, G. C. Partida, D. Dhamija, S. Das, S. L. Elson, T. Filshtein, K. Fletez-Brant, P. Fontanillas, W. Freyman, P. M. Gandhi, B. Hicks, D. A. Hinds, K. E. Huber, E. M. Jewett, Y. Jiang, A. Kleinman, K. Kukar, V. Lane, K.-H. Lin, M. Lowe, M. K. Luff, J. C. McCreight, M. H. McIntyre, K. F. McManus, S. J. Micheletti, M. E. Moreno, J. L. Mountain, S. V. Mozaffari,

P. Nandakumar, E. S. Noblin, J. O'Connell, A. A. Petrakovitz, G. D. Poznik, M. Schumacher, A. J. Shastri, J. F. Shelton, J. Shi, S. Shringarpure, C. Tian, V. Tran, J. Y. Tung, X. Wang, W. Wang, C. H. Weldon, P. Wilton, J. M. Sealock, L. K. Davis, N. L. Pedersen, C. A. Reynolds, I. K. Karlsson, S. Magnusson, H. Stefansson, S. Thordardottir, P. V. Jonsson, J. Snaedal, A. Zettergren, I. Skoog, S. Kern, M. Waern, H. Zetterberg, K. Blennow, E. Stordal, K. Hveem, J.-A. Zwart, L. Athanasiu, P. Selnes, I. Saltvedt, S. B. Sando, I. Ulstein, S. Djurovic, T. Fladby, D. Aarsland, G. Selbæk, S. Ripke, K. Stefansson, O. A. Andreassen, and D. Posthuma, "A genome-wide association study with 1,126,563 individuals identifies new risk loci for alzheimer's disease," *Nature Genetics*, vol. 53, pp. 1276–1282, 9 2021.

39. D. Demontis, R. K. Walters, J. Martin, M. Mattheisen, T. D. Als, E. Agerbo, G. Baldursson, R. Belliveau, J. Bybjerg-Grauholm, M. Bækvad-Hansen, *et al.*, "Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder," *Nature Genetics*, vol. 51, no. 1, pp. 63–75, 2019.

40. H. J. Watson, Z. Yilmaz, L. M. Thornton, C. Hübel, J. R. Coleman, H. A. Gaspar, J. Bryois, A. Hinney, V. M. Leppä, M. Mattheisen, *et al.*, "Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa," *Nature Genetics*, vol. 51, no. 8, pp. 1207–1214, 2019.

41. J. Grove, S. Ripke, T. D. Als, M. Mattheisen, R. K. Walters, H. Won, J. Pallesen, E. Agerbo, O. A. Andreassen, R. Anney, *et al.*, "Identification of common genetic risk variants for autism spectrum disorder," *Nature Genetics*, vol. 51, no. 3, pp. 431–444, 2019.

42. N. Mullins, A. J. Forstner, K. S. O'Connell, B. Coombes, J. R. I. Coleman, Z. Qiao, T. D. Als, T. B. Bigdeli, S. Børte, J. Bryois, A. W. Charney, O. K. Drange, M. J. Gandal, S. P. Hagenaars, M. Ikeda, N. Kamitaki, M. Kim, K. Krebs, G. Panagiotaropoulou, B. M. Schilder, L. G. Sloofman, S. Steinberg, V. Trubetskoy, B. S. Winsvold, H.-H. Won, L. Abramova, K. Adorjan, E. Agerbo, M. A. Eissa, D. Albani, N. Alliey-Rodriguez, A. Anjorin, V. Antilla, A. Antoniou, S. Awasthi, J. H. Baek, M. Bækvad-Hansen, N. Bass, M. Bauer, E. C. Beins, S. E. Bergen, A. Birner, C. B. Pedersen, E. Bøen, M. P. Boks, R. Bosch, M. Brum, B. M. Brumpton, N. Brunkhorst-Kanaan, M. Budde, J. Bybjerg-Grauholm, W. Byerley, M. Cairns, M. Casas, P. Cervantes, T.-K. Clarke, C. Cruceanu, A. Cuellar-Barboza, J. Cunningham, D. Curtis, P. M. Czerski, A. M. Dale, N. Dalkner, F. S. David, F. Degenhardt, S. Djurovic, A. L. Dobbyn, A. Douzenis, T. Elvsåshagen, V. Escott-Price, I. N. Ferrier, A. Fiorentino, T. M. Foroud, L. Forty, J. Frank, O. Frei, N. B. Freimer, L. Frisén, K. Gade, J. Garnham, J. Gelernter, M. G. Pedersen, I. R. Gizer, S. D. Gordon, K. Gordon-Smith, T. A. Greenwood, J. Grove, J. Guzman-Parra, K. Ha, M. Haraldsson, M. Hautzinger, U. Heilbronner, D. Hellgren, S. Herms, P. Hoffmann, P. A. Holmans, L. Huckins, S. Jamain, J. S. Johnson, J. L. Kalman, Y. Kamatani, J. L. Kennedy, S. Kittel-Schneider, J. A. Knowles, M. Kogevinas, M. Koromina, T. M. Kranz, H. R. Kranzler, M. Kubo, R. Kupka, S. A. Kushner, C. Lavebratt, J. Lawrence, M. Leber, H.-J. Lee, P. H. Lee, S. E. Levy, C. Lewis, C. Liao, S. Lucae, M. Lundberg, D. J. MacIntyre, S. H. Magnusson, W. Maier, A. Maihofer, D. Malaspina, E. Maratou, L. Martinsson, M. Mattheisen, S. A. McCarroll, N. W. McGregor, P. McGuffin, J. D. McKay, H. Medeiros, S. E. Medland, V. Millischer, G. W. Montgomery, J. L. Moran, D. W. Morris, T. W. Mühleisen, N. O'Brien, C. O'Donovan, L. M. O. Loohuis, L. Oruc, S. Papiol, A. F. Pardiñas, A. Perry, A. Pfennig, E. Porichi, J. B. Potash, D. Quested, T. Raj, M. H. Rapaport, J. R. DePaulo, E. J. Regeer, J. P. Rice, F. Rivas, M. Rivera, J. Roth, P. Roussos, D. M. Ruderfer, C. Sánchez-Mora, E. C. Schulte, F. Senner, S. Sharp, P. D. Shilling, E. Sigurdsson, L. Sirignano, C. Slaney, O. B. Smeland, D. J. Smith, J. L. Sobell, C. S. Hansen, M. S. Artigas, A. T. Spijker, D. J. Stein, J. S. Strauss, B. Światkowska, C. Terao, T. E. Thorgeirsson, C. Toma, P. Tooney, E.-E. Tsermpini, M. P. Vawter, H. Vedder, J. T. R. Walters, S. H. Witt, S. Xi, W. Xu, J. M. K. Yang, A. H. Young, H. Young, P. P. Zandi, H. Zhou, L. Zillich, R. Adolfsson, I. Agartz, M. Alda, L. Alfredsson, G. Babadjanova, L. Backlund, B. T. Baune, F. Bellivier, S. Bengesser, W. H. Berrettini, D. H. R. Blackwood, M. Boehnke, A. D. Børglum, G. Breen, V. J. Carr, S. Catts, A. Corvin, N. Craddock, U. Dannlowski, D. Dikeos, T. Esko, B. Etain, P. Ferentinos, M. Frye, J. M. Fullerton, M. Gawlik, E. S. Gershon, F. S. Goes, M. J. Green, M. Grigoroiu-Serbanescu, J. Hauser, F. Henskens, J. Hillert, K. S. Hong, D. M. Hougaard, C. M. Hultman, K. Hveem, N. Iwata, A. V. Jablensky, I. Jones, L. A. Jones, R. S. Kahn, J. R. Kelsoe, G. Kirov, M. Landén, M. Leboyer, C. M. Lewis, Q. S. Li, J. Lissowska, C. Lochner, C. Loughland, N. G. Martin, C. A. Mathews, F. Mayoral, S. L. McElroy, A. M. McIntosh, F. J. McMahon, I. Melle, P. Michie, L. Milani, P. B. Mitchell, G. Morken, O. Mors, P. B. Mortensen, B. Mowry, B. Müller-Myhsok, R. M. Myers, B. M. Neale, C. M. Nievergelt, M. Nordentoft, M. M. Nöthen, M. C. O'Donovan, K. J. Oedegaard, T. Olsson, M. J. Owen, S. A. Paciga, C. Pantelis, C. Pato, M. T. Pato, G. P. Patrinos, R. H. Perlis, D. Posthuma, J. A. Ramos-Quiroga, A. Reif, E. Z. Reininghaus, M. Ribasés, M. Rietschel, S. Ripke, G. A. Rouleau, T. Saito, U. Schall, M. Schalling, P. R. Schofield, T. G. Schulze, L. J. Scott, R. J. Scott, A. Serretti, C. S. Weickert, J. W. Smoller, H. Stefansson, K. Stefansson, E. Stordal, F. Streit, P. F. Sullivan, G. Turecki, A. E. Vaaler, E. Vieta, J. B. Vincent, I. D. Waldman, T. W. Weickert, T. Werge, N. R. Wray, J.-A. Zwart, J. M. Biernacka, J. I. Nurnberger, S. Cichon, H. J. Edenberg, E. A. Stahl, A. McQuillin, A. D. Florio, R. A. Ophoff, and O. A. Andreassen, "Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology," *Nature Genetics*, vol. 53, pp. 817–829, 6 2021.

43. D. M. Howard, M. J. Adams, T.-K. Clarke, J. D. Hafferty, J. Gibson, M. Shirali, J. R. Coleman, S. P. Hagenaars, J. Ward, E. M. Wigmore, *et al.*, "Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions," *Nature Neuroscience*, vol. 22, no. 3, pp. 343–352, 2019.

44. C. M. Nievergelt, A. X. Maihofer, T. Klengel, E. G. Atkinson, C.-Y. Chen, K. W. Choi, J. R. I. Coleman, S. Dalvie, L. E. Duncan, J. Gelernter, D. F. Levey, M. W. Logue, R. Polimanti, A. C. Provost, A. Ratanatharathorn, M. B. Stein, K. Torres, A. E. Aiello, L. M. Almli, A. B. Amstadter, S. B. Andersen, O. A. Andreassen, P. A. Arbisi, A. E. Ashley-Koch, S. B. Austin, E. Avdibegovic, D. Babić, M. Bækvad-Hansen, D. G. Baker, J. C. Beckham, L. J. Bierut, J. I. Bisson, M. P. Boks, E. A. Bolger, A. D. Børglum, B. Bradley, M. Brashear, G. Breen, R. A. Bryant, A. C. Bustamante, J. Bybjerg-Grauholm, J. R. Calabrese, J. M. C. de Almeida, A. M. Dale, M. J. Daly, N. P. Daskalakis, J. Deckert, D. L. Delahanty, M. F. Dennis, S. G. Disner, K. Domschke, A. Dzubur-Kulenovic, C. R. Erbes, A. Evans, L. A. Farrer, N. C. Feeny, J. D. Flory, D. Forbes, C. E. Franz, S. Galea, M. E. Garrett, B. Gelaye, E. Geuze, C. Gillespie, A. G. Uka, S. D. Gordon, G. Guffanti, R. Hammamieh, S. Harnal, M. A. Hauser, A. C. Heath, S. M. J. Hemmings, D. M. Hougaard, M. Jakovljevic, M. Jett, E. O. Johnson, I. Jones, T. Jovanovic, X.-J. Qin, A. G. Junglen, K.-I. Karstoft, M. L. Kaufman, R. C. Kessler, A. Khan, N. A. Kimbrel, A. P. King, N. Koen, H. R. Kranzler, W. S. Kremen, B. R. Lawford, L. A. M. Lebois, C. E. Lewis, S. D. Linnstaedt, A. Lori, B. Lugonja, J. J. Luykx, M. J. Lyons, J. Maples-Keller, C. Marmar, A. R. Martin, N. G. Martin, D. Maurer, M. R. Mavissakalian, A. McFarlane, R. E. McGlinchey, K. A. McLaughlin, S. A. McLean, S. McLeay, D. Mehta, W. P. Milberg, M. W. Miller, R. A. Morey, C. P. Morris, O. Mors, P. B. Mortensen, B. M. Neale, E. C. Nelson, M. Nordentoft, S. B. Norman, M. O'Donnell, H. K. Orcutt, M. S. Panizzon, E. S. Peters, A. L. Peterson, M. Peverill, R. H. Pietrzak, M. A. Polusny, J. P. Rice, S. Ripke, V. B. Risbrough, A. L. Roberts, A. O. Rothbaum, B. O. Rothbaum, P. Roy-Byrne, K. Ruggiero, A. Rung, B. P. F. Rutten, N. L. Saccone, S. E. Sanchez, D. Schijven, S. Seedat, A. V. Seligowski, J. S. Seng, C. M. Sheerin, D. Silove, A. K. Smith, J. W. Smoller, S. R. Sponheim, D. J. Stein, J. S. Stevens, J. A. Sumner, M. H. Teicher, W. K. Thompson, E. Trapido, M. Uddin, R. J. Ursano, L. L. van den Heuvel, M. V. Hooff, E. Vermetten, C. H. Vinkers, J. Voisey, Y. Wang, Z. Wang, T. Werge, M. A. Williams, D. E. Williamson, S. Winternitz, C. Wolf, E. J. Wolf, J. D. Wolff, R. Yehuda, R. M. Young, K. A. Young, H. Zhao, L. A. Zoellner, I. Liberzon, K. J. Ressler, M. Haas, and K. C. Koenen, "International meta-analysis of ptsd genome-wide association studies identifies sex- and ancestry-specific genetic risk loci," *Nature Communications*, vol. 10, p. 4558, 10 2019.

45. S. Ripke, J. T. Walters, M. C. O'Donovan, S. W. G. of the Psychiatric Genomics Consortium, *et al.*, "Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia," *MedRxiv*, 2020.

46. J. J. Lee, R. Wedow, A. Okbay, E. Kong, O. Maghzian, M. Zacher, T. A. Nguyen-Viet, P. Bowers, J. Sidorenko, R. K. Linnér, *et al.*, "Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals," *Nature genetics*, vol. 50, pp. 1112–1121, 2018.

47. A. Okbay, Y. Wu, N. Wang, H. Jayashankar, M. Bennett, S. M. Nehzati, J. Sidorenko, H. Kweon, G. Goldman, T. Gjorgjieva, Y. Jiang, B. Hicks, C. Tian, D. A. Hinds, R. Ahlskog, P. K. E. Magnusson, S. Oskarsson, C. Hayward, A. Campbell, D. J. Porteous, J. Freese, P. Herd, M. Agee, B. Alipanahi, A. Auton, R. K. Bell, K. Bryc, S. L. Elson, P. Fontanillas, N. A. Furlotte, D. A. Hinds, K. E. Huber, A. Kleinman, N. K. Litterman, J. C. McCreight, M. H. McIntyre, J. L. Mountain, C. A. M. Northover, S. J. Pitts, J. F. Sathirapongsasuti, O. V. Sazonova, J. F. Shelton, S. Shringarpure, J. Y. Tung, V. Vacic, C. H. Wilson, M. A. Fontana, T. H. Pers, C. A. Rietveld, G.-B. Chen, V. Emilsson, S. F. W. Meddens, J. K. Pickrell, K. Thom, P. Timshel, R. de Vlaming, A. Abdellaoui, T. S. Ahluwalia, J. Bacelis, C. Baumbach, G. Bjornsdottir, J. H. Brandsma, M. P. Concas, J. Derringer, T. E. Galesloot, G. Girotto, R. Gupta, L. M. Hall, S. E. Harris, E. Hofer, M. Horikoshi, J. E. Huffman, K. Kaasik, I. P. Kalafati, R. Karlsson, J. Lahti, S. J. van der Lee, C. de Leeuw, P. A. Lind, K.-O. Lindgren, T. Liu, M. Mangino, J. Marten, E. Mihailov, M. B. Miller, P. J. van der Most, C. Oldmeadow, A. Payton, N. Pervjakova, W. J. Peyrot, Y. Qian, O. Raitakari, R. Rueedi, E. Salvi, B. Schmidt, K. E. Schraut, J. Shi, A. V. Smith, R. A. Poot, B. S. Pourcain, A. Teumer, G. Thorleifsson, N. Verweij, D. Vuckovic, J. Wellmann, H.-J. Westra, J. Yang, W. Zhao, Z. Zhu, B. Z. Alizadeh, N. Amin, A. Bakshi, S. E. Baumeister, G. Biino, K. Bønnelykke, P. A. Boyle, H. Campbell, F. P. Cappuccio, G. Davies, J.-E. D. Neve, P. Deloukas, I. Demuth, J. Ding, P. Eibich, L. Eisele, N. Eklund, D. M. Evans, J. D. Faul, M. F. Feitosa, A. J. Forstner, I. Gandin, B. Gunnarsson, B. V. Halldórsson, T. B. Harris, A. C. Heath, L. J. Hocking, E. G. Holliday, G. Homuth, M. A. Horan, J.-J. Hottenga, P. L. de Jager, P. K. Joshi, A. Jugessur, M. A. Kaakinen, M. Kähönen, S. Kanoni, L. Keltigangas-Järvinen, L. A. L. M. Kiemeney, I. Kolcic, S. Koskinen, A. T. Kraja, M. Kroh, Z. Kutalik, A. Latvala, L. J. Launer, M. P. Lebreton, D. F. Levinson, P. Lichtenstein, P. Lichtner, D. C. M. Liewald, A. Loukola, P. A. Madden, R. Mägi, T. Mäki-Opas, R. E. Marioni, P. Marques-Vidal, G. A. Meddens, G. McMahon, C. Meisinger, T. Meitinger, Y. Milaneschi, L. Milani, G. W. Montgomery, R. Myhre, C. P. Nelson, D. R. Nyholt, W. E. R. Ollier, A. Palotie, L. Paternoster, N. L. Pedersen, K. E. Petrovic, K. Räikkönen, S. M. Ring, A. Robino, O. Rostapshova, I. Rudan, A. Rustichini, V. Salomaa, A. R. Sanders, A.-P. Sarin, H. Schmidt, R. J. Scott, B. H. Smith, J. A. Smith, J. A. Staessen, E. Steinhagen-Thiessen, K. Strauch, A. Terracciano, M. D. Tobin, S. Ulivi, S. Vaccargiu, L. Quaye, F. J. A. van Rooij, C. Venturini, A. A. E. Vinkhuyzen, U. Völker, H. Völzke, J. M. Vonk, D. Vozzi, J. Waage, E. B. Ware, G. Willemsen, J. R. Attia, D. A. Bennett, K. Berger, L. Bertram, H. Bisgaard, D. I. Boomsma, I. B. Borecki, U. Bültmann, C. F. Chabris, F. Cucca, D. Cusi, I. J. Deary, G. V. Dedoussis, C. M. van Duijn, J. G. Eriksson, B. Franke, L. Franke, P. Gasparini, P. V. Gejman, C. Gieger, H.-J. Grabe, J. Gratten, P. J. F. Groenen, V. Gudnason, P. van der Harst, W. Hoffmann, E. Hyppönen, W. G. Iacono, B. Jacobsson, M.-R. Järvelin, K.-H. Jöckel, J. Kaprio, S. L. R. Kardia, T. Lehtimäki, S. F. Lehrer, N. G. Martin, M. McGue, A. Metspalu, N. Pendleton, B. W. J. H. Penninx, M. Perola, N. Pirastu, M. Pirastu, O. Polasek, D. Posthuma, C. Power, M. A. Province, N. J. Samani, D. Schlessinger, R. Schmidt, T. I. A. Sørensen, T. D. Spector, K. Stefansson, U. Thorsteinsdottir, A. R. Thurik, N. J. Timpson, H. Tiemeier, A. G. Uitterlinden, V. Vitart, P. Vollenweider, D. R. Weir, J. F. Wilson, A. F. Wright, D. C. Conley, R. F. Krueger, G. D. Smith, A. Hofman, D. I. Laibson, S. E. Medland, J. Yang, T. Esko, C. Watson, J. Jala, D. Conley, P. D. Koellinger, M. Johannesson, D. Laibson, M. N. Meyer, J. J. Lee, A. Kong, L. Yengo, D. Cesarini, P. Turley, P. M. Visscher, J. P. Beauchamp, D. J. Benjamin, and A. I. Young, "Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals," *Nature Genetics*, vol. 54, pp. 437–449, 4 2022.

48. R. K. Linnér, P. Biroli, E. Kong, S. F. W. Meddens, R. Wedow, M. A. Fontana, M. Lebreton, S. P. Tino, A. Abdellaoui, A. R. Hammerschlag, M. G. Nivard, A. Okbay, C. A. Rietveld, P. N. Timshel, M. Trzaskowski, R. de Vlaming, C. L. Zünd, Y. Bao, L. Buzdugan, A. H. Caplin, C.-Y. Chen, P. Eibich, P. Fontanillas, J. R. Gonzalez, P. K. Joshi, V. Karhunen, A. Kleinman, R. Z. Levin, C. M. Lill, G. A. Meddens, G. Muntané, S. Sanchez-Roige, F. J. van Rooij, E. Taskesen, Y. Wu, F. Zhang, A. Auton, J. D. Boardman, D. W. Clark, A. Conlin, C. C. Dolan, U. Fischbacher, P. J. F. Groenen, K. M. Harris, G. Hasler, A. Hofman, M. A. Ikram, S. Jain, R. Karlsson, R. C. Kessler, M. Kooyman, J. MacKillop, M. Männikkö, C. Morcillo-Suarez, M. B. McQueen, K. M. Schmidt, M. C. Smart, M. Sutter, A. R. Thurik, A. G. Uitterlinden, J. White, H. de Wit, J. Yang, L. Bertram, D. I. Boomsma, T. Esko, E. Fehr, D. A. Hinds, M. Johannesson, M. Kumari, D. Laibson, P. K. E. Magnusson, M. N. Meyer, A. Navarro, A. A. Palmer, T. H. Pers, D. Posthuma, D. Schunk, M. B. Stein, R. Svento, H. Tiemeier, P. R. H. J. Timmers, P. Turley, R. J. Ursano, G. G. Wagner, J. F. Wilson, J. Gratten, J. J. Lee, D. Cesarini, D. J. Benjamin, P. D. Koellinger, and J. P. Beauchamp, "Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences," *Nature Genetics*, vol. 51, pp. 245–257, 2 2019.

49. R. C. Team, "R: A language and environment for statistical computing," 2013.