# Transcriptome-Based Prediction of Methylphenidate Efficacy: Implications for Precision Medicine in Psychiatric Disorders

# Exploring the Genetic Landscape of Methylphenidate Response: Insights into Treatment Optimization

Muhammad Elsadany1,2, Taylor R. Thomas1,2, Lucas G. Casten1,2, Jacob J. Michaelson1,2,3,4,5 \*

**1** Department of Psychiatry, University of Iowa, Iowa City, IA

**2** Interdisciplinary Graduate Program in Genetics, University of Iowa, Iowa City, IA

**3** Department of Communication Sciences & Disorders, University of Iowa, Iowa City, IA

**4** Iowa Neuroscience Institute, University of Iowa, Iowa City, IA

**5** Department of Biomedical Engineering, University of Iowa, Iowa City, IA

**\*** jacob-michaelson@uiowa.edu

# 1. Abstract

**Background:** The variability in drug response among individuals with neurodevelopmental disorders, such as attention deficit hyperactivity disorder (ADHD), poses a significant challenge for effective pharmacological management. This study aimed to develop a novel method for predicting methylphenidate (MPH) response using genetic profiles and transcriptomic imputation and validate its predictive capacity using data from the SPARK and ABCD cohorts.

**Methods:** A multi-step pipeline was employed, involving the imputation of transcriptomic profiles for excitatory neurons using eQTL weights and comparison to the MPH transcriptomic signature derived from the Connectivity Map (CMap) database. The predicted drug response was then validated against objective measures of attention and cognitive performance.

**Results:** In the SPARK cohort, significant correlations were observed between reported MPH effectiveness and communication abilities, with positive correlations for improved communication and negative correlations for odd behaviors and deliberate self-injury. In the ABCD cohort, significant correlations were found between archetypes derived from the stop signal fMRI task (SST) and performance measures from the NIH Toolbox. Regression models revealed significant positive main effects for our predicted MPH response in predicting better performance in SST archetype 2. A significant positive effect was observed for the interaction term in predicting the same archetype, suggesting better performance for individuals predicted to respond to MPH but taking it on the day of the task. Similar patterns were observed in the regression models predicting performance in the NIH Toolbox.

**Conclusion:** This study presents a novel method for predicting MPH response using genetic profiles and transcriptomic imputation. By incorporating objective measures and leveraging genetic information, this approach offers a promising avenue for personalized treatment strategies in neurodevelopmental disorders. The findings contribute to a deeper understanding of the genetic underpinnings of drug response and lay the foundation for optimizing medication choices and dosages to enhance treatment outcomes.

**Keywords:** ADHD, methylphenidate response, genetic profiles, transcriptomic imputation, polygenic scores, personalized medicine.

# 2. Introduction

Although progress in human genetics has drastically accelerated in the decades since the Human Genome Project, the current impact of genomic data on treatments for disease is limited. Still, individual genotype profiles have been known to be a major player in differentiating response to different pharmacological compounds (Dunnenberger et al., 2015; Relling & Evans, 2015). The field of pharmacogenomics is rapidly growing, with the aim of providing personalized interventions for patients by uncovering the genetic basis of inter-individual variability in drug response. However, currently available pharmacogenomic services are limited in their ascertainment of genetic variation, with an emphasis on drug metabolism, known drug targets, and known mechanisms of action. This approach may result in blind spots relating to off-target effects and additional/alternative therapeutic mechanisms. A more comprehensive understanding of drug response could provide recommendations that maximize therapeutic effect while minimizing side effects.

Attention deficit hyperactivity disorder (ADHD) is a prevalent condition, affecting approximately 6 million (9.8%) of US children aged 3–17 years ever diagnosed with ADHD (using data from 2016-2019), with significant variations across different populations and cultures (CDC, 2022). The etiology of ADHD is complex, involving a combination of genetic, environmental, and neurobiological factors, and its diagnosis often occurs alongside comorbid conditions such as oppositional defiant disorder (ODD), conduct disorder (CD), and learning disabilities (LD) (Pliszka, 1998). While several treatment options are available, methylphenidate remains one of the most prescribed medications due to its efficacy in alleviating ADHD symptoms. However, predicting individual response to methylphenidate treatment poses a challenge, necessitating a comprehensive exploration of various factors that influence treatment outcomes. This paper provides an updated review of ADHD in children, its comorbidities, and contemporary perspectives on predicting response to methylphenidate, integrating relevant research and current statistics.

Individuals with ADHD often require pharmacological interventions to manage their symptoms and improve their daily functioning (American-Academy-of-Pediatrics, 2022). Methylphenidate, a commonly prescribed medication for ADHD, has shown efficacy in enhancing attention and reducing hyperactivity in many individuals (Storebo et al., 2018). However, there is significant variability in individual drug response, with some individuals experiencing substantial benefits while others show limited improvement or even adverse effects (Cortese et al., 2018; de Faria et al., 2022; Kimko et al., 1999). Understanding the factors that contribute to this variability in drug response is crucial for optimizing treatment strategies and achieving personalized medicine approaches.

Traditionally, drug response assessments have relied on subjective reporting from patients, parents, or caregivers (Chang et al., 2021; Grazioli et al., 2021; Vertessen et al., 2023). Although valuable, these assessments are inherently subjective and can be influenced by various factors, including placebo effects and biases. Objective measures, such as neuroimaging or behavioral tasks, have also been employed to evaluate drug effectiveness. However, these measures often require the patient to be on medication for several weeks before any measurable changes can be observed, making them time-consuming and less suitable for rapid treatment optimization.

To address these limitations, there is a growing interest in leveraging genetic information to predict drug response. Genetic factors play a significant role in individual variability in drug metabolism, receptor binding, and overall pharmacodynamics (Belle & Singh, 2008; Evans & Johnson, 2001; Ingelman-Sundberg et al., 2018; Roden et al., 2011; Scharfe et al., 2017; Weinshilboum, 2003; Wilke & Dolan, 2011). Polygenic scores (PGS) derived from genome-wide association studies (GWAS) have been used to capture the genetic contribution to complex traits and diseases. By incorporating PGS into predictive models, it is possible to estimate an individual's likelihood of responding favorably to a specific medication.

In this study, we aimed to develop a method for predicting drug response based on genetic profiles and apply it to methylphenidate, a widely used medication for ADHD. We hypothesized that by leveraging genetic information and transcriptomic profiles, we could identify individuals who are more likely to respond positively to methylphenidate, thereby enabling personalized treatment strategies. Our approach involved imputing transcriptomic profiles for excitatory neurons based on eQTL weights and comparing them to the methylphenidate transcriptomic signature derived from the Connectivity Map (CMap) database.

To validate the effectiveness of our predictive pipeline, we utilized data from the SPARK[[1]](#footnote-1) and ABCD[[2]](#footnote-2) cohorts. The SPARK cohort provided subjective reports on methylphenidate effectiveness and social communication abilities, while the ABCD cohort provided objective measures of attention and cognitive performance. By comparing our predicted drug response with the objective measures, we aimed to establish the utility of our approach in predicting clinical outcomes and identifying individuals who are more likely to benefit from methylphenidate treatment.

The findings from this study have the potential to significantly impact clinical practice by enabling more targeted and individualized treatment strategies for neurodevelopmental disorders. By accurately predicting drug response, clinicians can optimize medication choices and dosages, reducing the trial-and-error approach and minimizing potential side effects. Moreover, our approach offers a valuable framework for exploring the genetic underpinnings of drug response, contributing to the broader understanding of the pharmacogenetics of neurodevelopmental disorders.

In summary, this study aimed to develop and validate a method for predicting methylphenidate response based on genetic profiles. By incorporating transcriptomic imputation and comparing it to the drug signature, we sought to provide an objective and personalized approach to drug response prediction. The results of this study have implications for improving treatment outcomes in neurodevelopmental disorders and advancing our understanding of the genetic factors influencing drug response.

# 3. Results

## 3.1. Correlation between MPH Effectiveness and Communication Abilities

We first examined the correlation between reported MPH effectiveness and the social communication abilities of the child as assessed by the Social Communication Questionnaire (SCQ). Our analysis revealed a significant correlation between the effectiveness of MPH and the child's communication abilities Figure 1-A. Specifically, we found a positive correlation between reported MPH effectiveness and the child's use of verbal communication, including the ability to use phrases. Conversely, we observed a negative correlation between reported MPH effectiveness and behaviors such as odd communication patterns or deliberate self-injury.

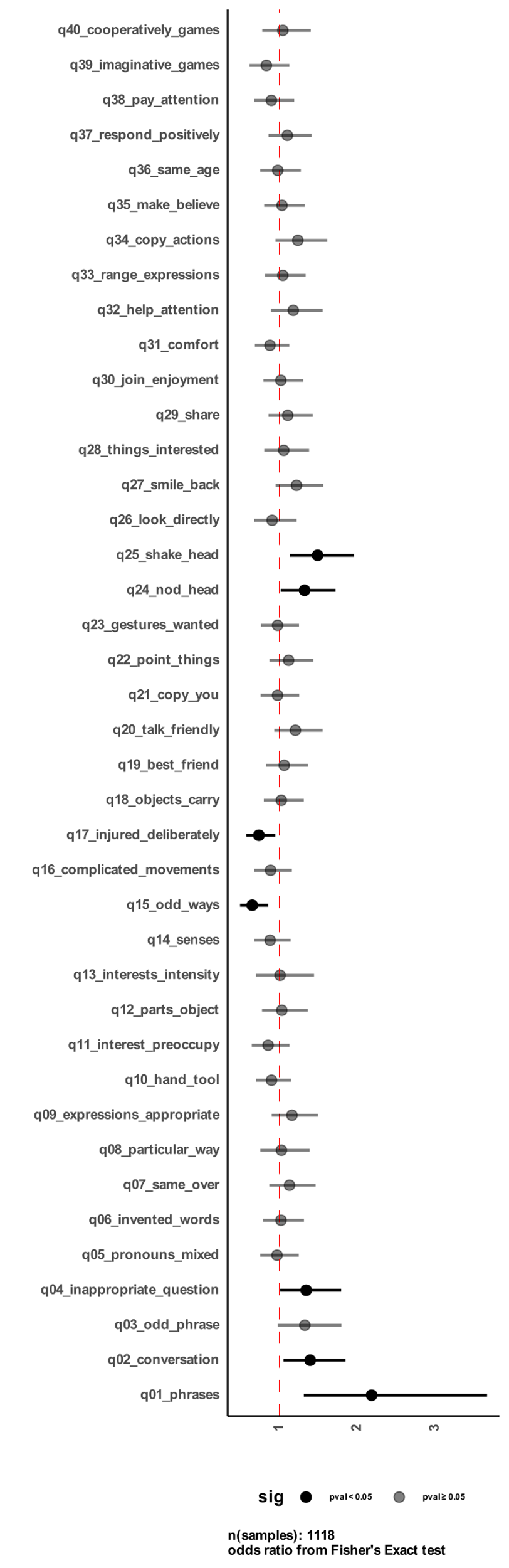
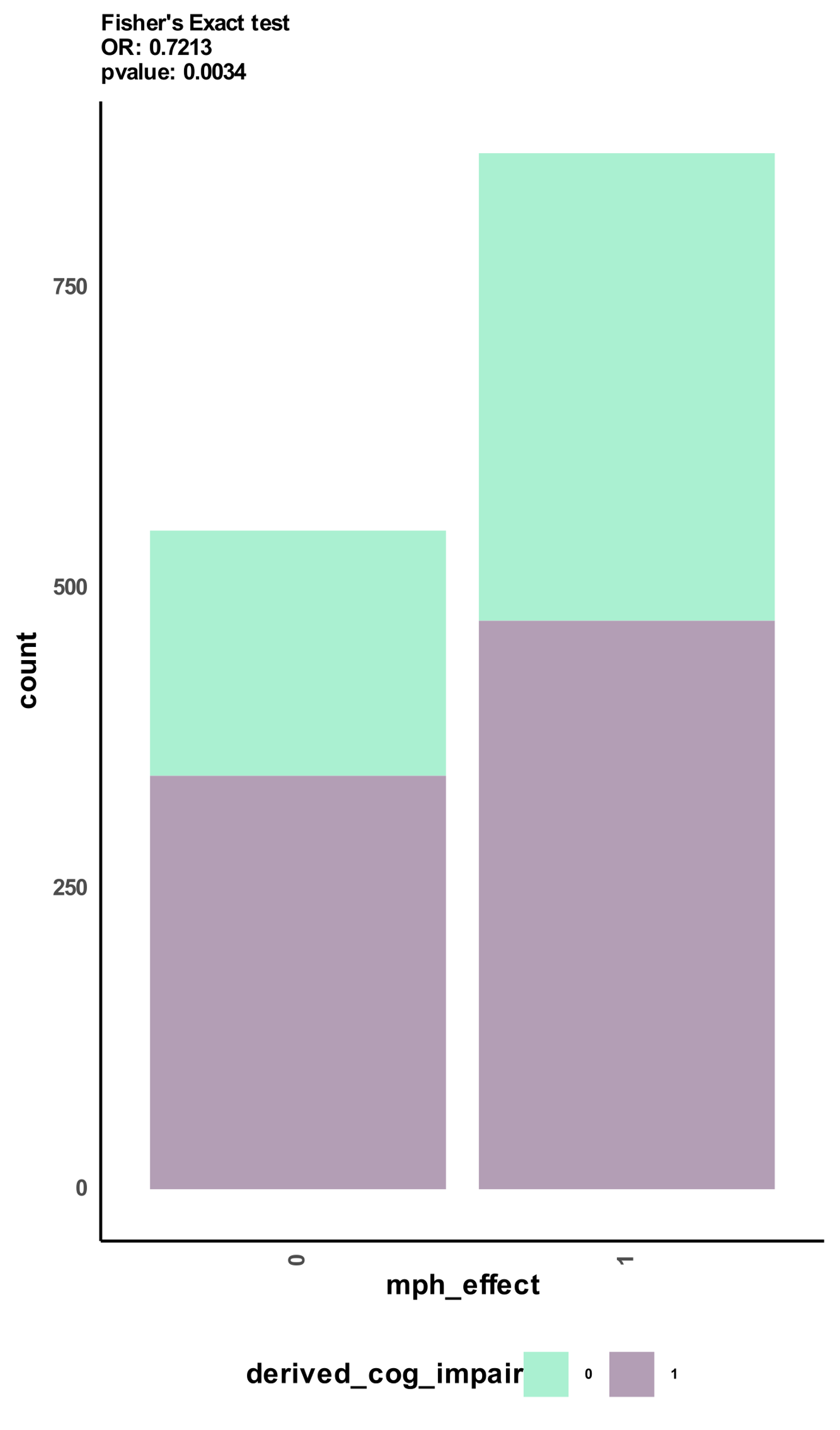
## 3.2. Correlation between MPH Effectiveness and Cognitive Impairment

To further investigate the relationship between MPH effectiveness and cognitive abilities, we analyzed data from SPARK participants using a binary variable indicating cognitive impairment. Our analysis revealed a negative correlation between reported MPH effectiveness and cognitive impairment (Odds Ratio = 0.7213; Fisher's Exact p-value = 0.0034) Figure 1-B. These findings suggest that individuals with cognitive impairment may experience lower effectiveness of MPH treatment.

## 3.3. Correlation between MPH effectiveness and Polygenic Scores

To explore the genetic basis underlying the observed relationship between MPH effectiveness and cognitive abilities, we calculated polygenic scores (PGS) for all participants, correcting for the first five genetic principal components. We then examined the correlation between these PGS and reported MPH effectiveness. Interestingly, we found a positive correlation between symbol digit matching PGS (source: UKB; see methods) and MPH effectiveness (Spearman correlation coefficient = 0.0512; nominal p-value = 0.0683) Figure 1-C. This suggests that individuals with higher polygenic score for symbol digit matching may respond more positively to MPH treatment. A similar positive correlation was found with most of other cognitive PGS, however, none of these passed the significance threshold of nominal p-value < 0.05.

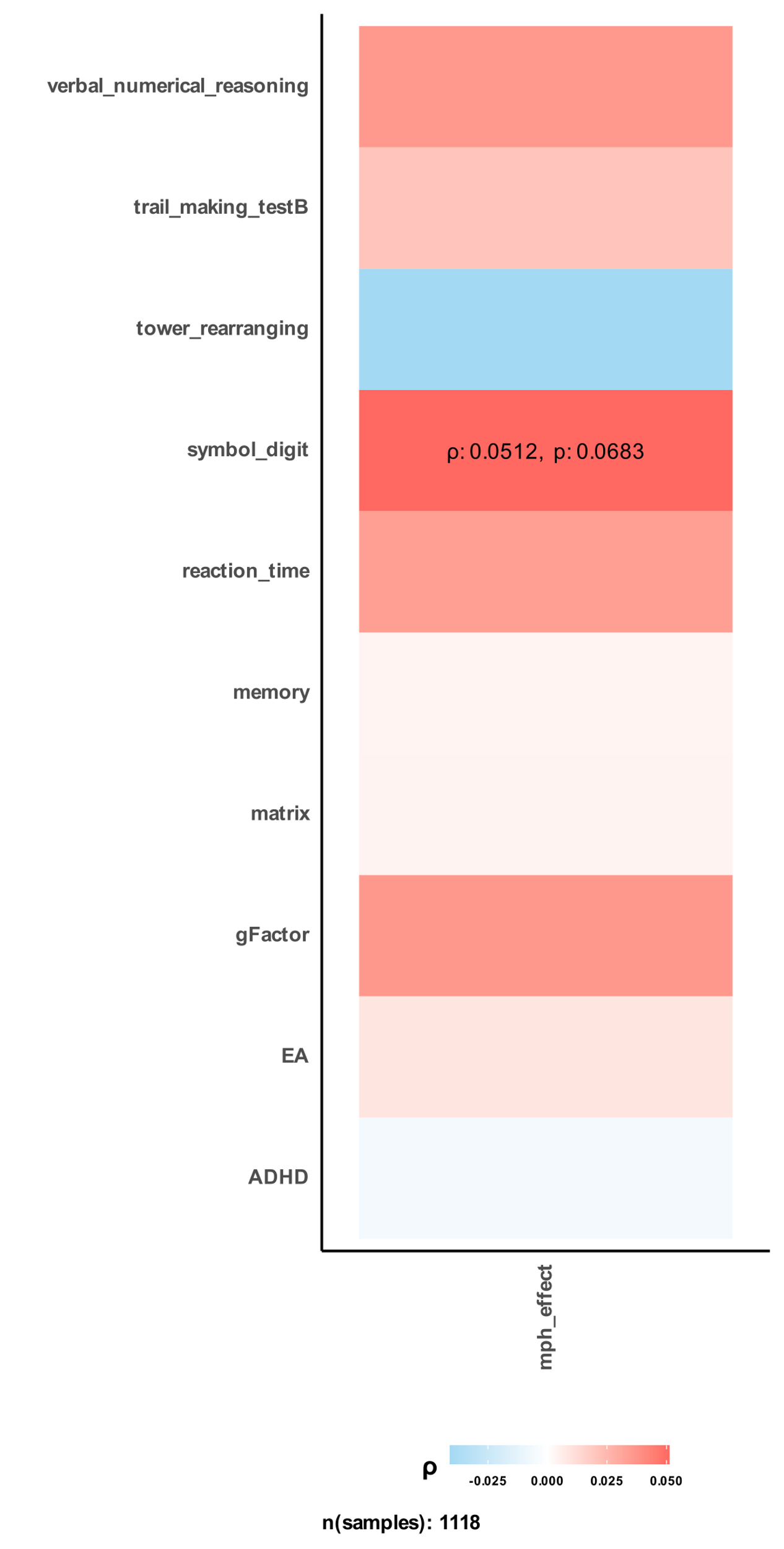
These results provide promising evidence linking MPH effectiveness to both communication abilities and cognitive functioning. However, further investigation is warranted to validate these findings and explore the underlying mechanisms in larger and more diverse cohorts.



C

B

A



**Figure 1 Correlation between MPH effectiveness and communication abilities, cognitive impairment, and polygenic scores in the SPARK cohort.**  
A) shows odds ratio (OR) between reported methylphenidate effectiveness (MPH\_effect) and different social communication questionnaire questions for 1118 samples from SPARK. B) correlation between cognitive impairment labels derived by SPARK and reported methylphenidate effectiveness by parents. C) shows Spearman rank-based correlation coefficients between MPH effectiveness and different calculated polygenic scores (PGS).

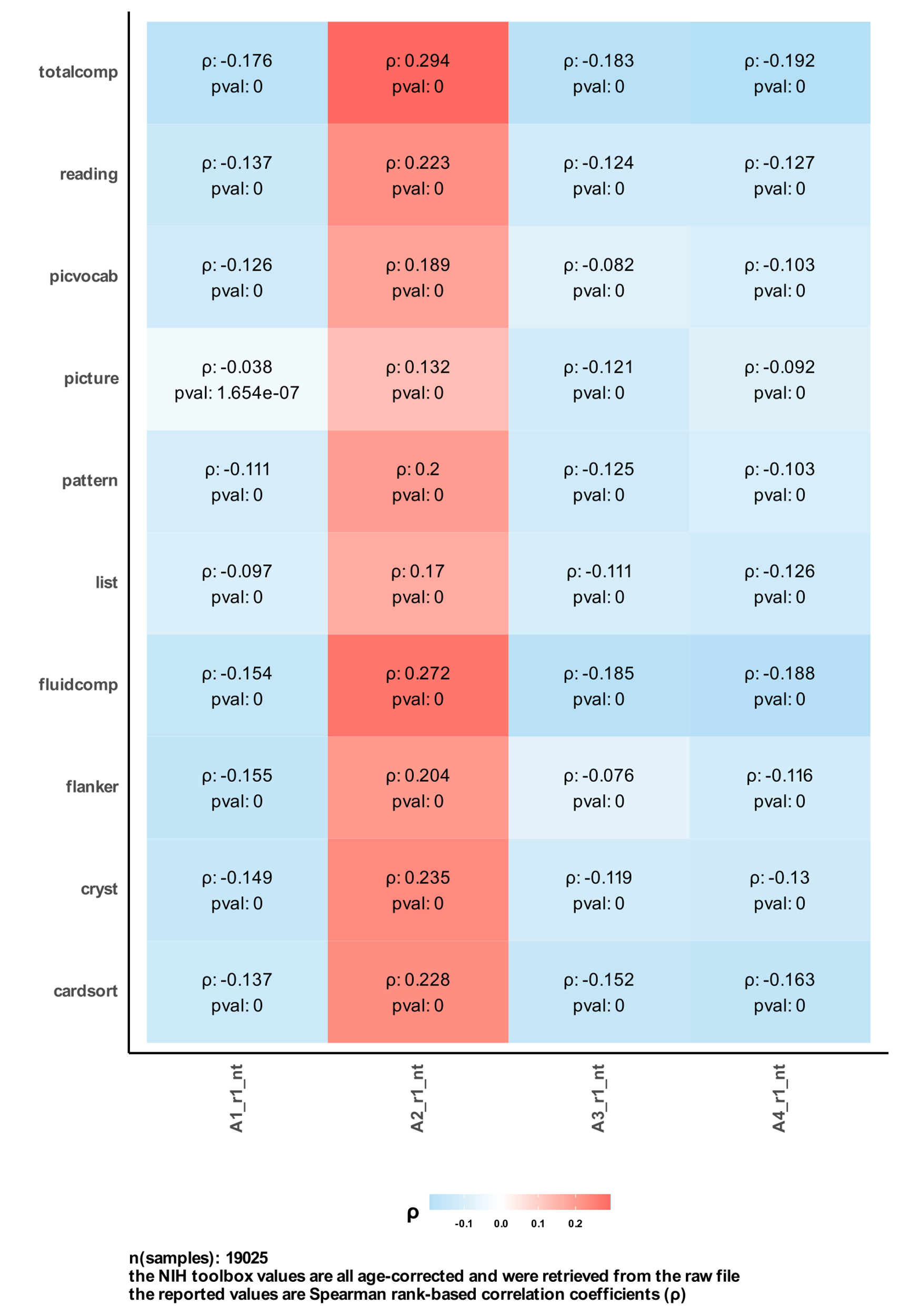
## 3.4. Validation of Genetic Component and Archetypal Analysis in the ABCD Cohort

To validate our findings regarding the genetic component predicting MPH effectiveness, we conducted a validation analysis using an independent sample from the Adolescent Brain Cognitive Development (ABCD) cohort. In this cohort, we utilized participants' performance in the first trial run of the stop signal fMRI task (SST) as an objective measure of attention.

We performed archetypal analysis on the SST measures to identify distinct patterns or archetypes representing different attention profiles (refer to the Methods section for details). To gain a better understanding of the meaning of these archetypes, we correlated their corresponding alphas with participants' performance in the NIH Toolbox measures. The NIH Toolbox provides a comprehensive assessment of various cognitive domains.

Our analysis revealed a significant correlation between the archetypes' alphas and performance in the NIH Toolbox measures Figure 2. Specifically, we observed that a higher alpha value in the second archetype (A2\_r1\_nt) was associated with better performance in the tasks, while higher alphas for the first, third, and fourth archetypes (A1\_r1\_nt, A3\_r1\_nt, and A4\_r1\_nt) were indicative of poorer performance. The opposite was also true, with lower alpha values corresponding to better performance for the latter three archetypes.

These findings strengthen the link between the archetypes derived from the SST and cognitive performance measured by the NIH Toolbox. This validation analysis provides additional support for the validity of our approach in identifying objective measures of attention and cognitive abilities, further supporting the genetic component associated with MPH effectiveness.



**Figure 2 Correlation between main SST archetypes and NIH Toolbox tasks.**   
The figure shows the Spearman correlation coefficients for 19025 ABCD participants between the main four archetypes of their performance in the stop signal fMRI task (SST) and their performance in the NIH Toolbox tasks.

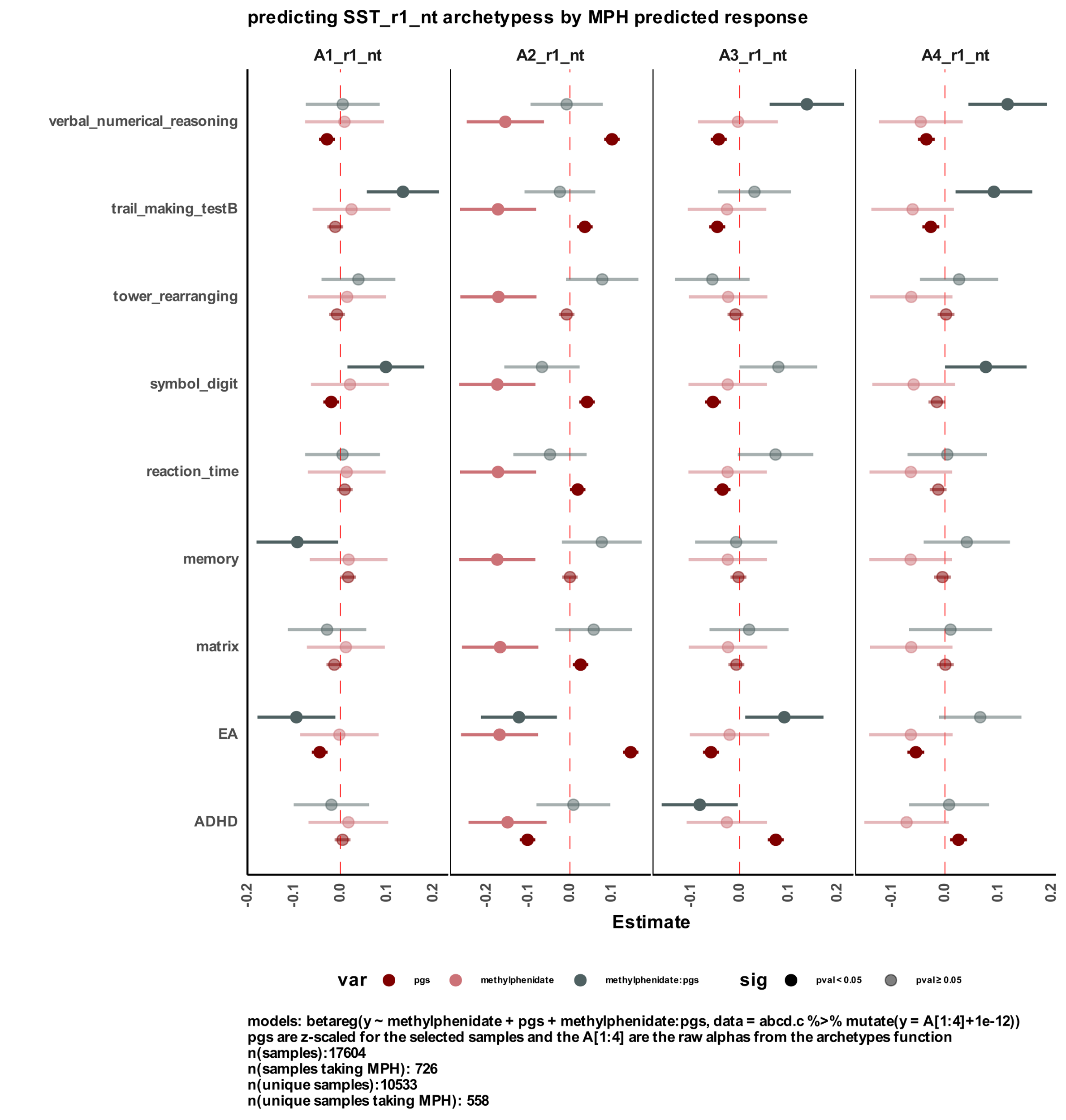
## 3.5. Association between Polygenic Scores and SST Archetypes

Furthermore, we examined the correlation between polygenic scores (PGS) and the four main archetypes derived from the stop signal fMRI task (SST) in the ABCD cohort. We employed a regression model to predict the alpha values of each archetype using the PGS as a main effect. The model also included two additional variables: one indicating whether participants were taking methylphenidate at the fMRI task day or not and another for the interaction term between the PGS and methylphenidate status. The estimates of each regression model are depicted in Figure 3, with separate models fitted for each PGS and archetype. Notably, for archetype 2 (A2\_r1\_nt), which is associated with better task performance, we observed a significant positive main effect for most of the cognitive PGS and education attainment (EA) PGS, indicating their predictive value in determining participants' performance.

~~It is worth mentioning that the main effect of methylphenidate was negatively correlated with participants performance indicated by archetype 2 (A2\_r1\_nt). From that, taking methylphenidate is correlated with poor performance in the task. We also noticed that the interaction term between each PGS and methylphenidate status had a different effect among archetypes, which is not expected. The three archetypes of A1\_r1\_nt, A3\_r1\_nt, and A4\_r1\_nt were shown to go in the same direction based on results in Figure 2, however, they seem to be different in terms of interaction with PGS (see correlation of EA with second and third archetype).~~

In our analysis, we observed intriguing patterns regarding the main effect of methylphenidate and its correlation with participants' performance, specifically as indicated by archetype 2 (A2\_r1\_nt). Surprisingly, the main effect of methylphenidate demonstrated a negative correlation with task performance, suggesting that the administration of methylphenidate was associated with poorer performance in the task.

Furthermore, we investigated the interaction between each polygenic score (PGS) and the status of methylphenidate usage, expecting consistent effects across archetypes. However, our findings revealed unexpected variations in the effects of the interaction term among different archetypes. While archetypes A1\_r1\_nt, A3\_r1\_nt, and A4\_r1\_nt exhibited similar patterns based on the results depicted in Figure 2, they displayed distinct interactions with PGS. Notably, the correlation of the interaction with EA PGS differed across the second and third archetypes, indicating potential heterogeneity in their underlying mechanisms. These intriguing observations warrant further investigation and highlight the complexity of the relationship between PGS, methylphenidate, and the diverse archetypes identified in our analysis.



**Figure 3 correlation between calculated polygenic scores and participants performance in SST.**  
Estimates for different polygenic scores to predict participants’ performance in the SST as summarized by the archetypes analysis. The regression model was fit as a beta regression with the formula Y ~ PGS + methylphenidate + PGS:methylphenidate. the methylphenidate represents a binary variable of taking methylphenidate or not on the day pf performing the SST.

## 3.6. Predicting Drug Response and its Impact on Performance Measures

To investigate the applicability of polygenic scores (PGS) in predicting drug response, we initially examined their performance in two independent cohorts: ABCD and SPARK. Surprisingly, we observed a significant discrepancy in the results obtained from these cohorts, with the PGS failing to demonstrate a consistent pattern as observed in the SPARK cohort. Given the limitations of PGS in accurately predicting drug response, we recognized the urgent need to develop a more robust and reliable tool that could extract the true genetic signals associated with drug response.

We developed a method for predicting drug response based solely on the genetic profiles of individuals and the specific drug under investigation. Using eQTL weights, we imputed transcriptomic profiles for excitatory neurons and compared them to the methylphenidate transcriptomic signature from the CMap database. The correlation between an individual's imputed transcriptome and the drug signature was used as a measure of drug effectiveness. A negative correlation indicated that the drug could neutralize the imputed transcriptome's effect, while a positive correlation suggested an exacerbation of gene expression in a similar direction to the imputed profile.

### 3.6.1. Performance in the SST

After predicting participants' methylphenidate response using this pipeline, we examined its impact on their performance in the SST. A beta regression model was employed to predict participants' archetype alpha weights, including the main effect of predicted response to methylphenidate, the status of taking methylphenidate during the fMRI task, and their interaction. The estimates for each term in the model by archetype revealed significant correlations Figure 4-A.

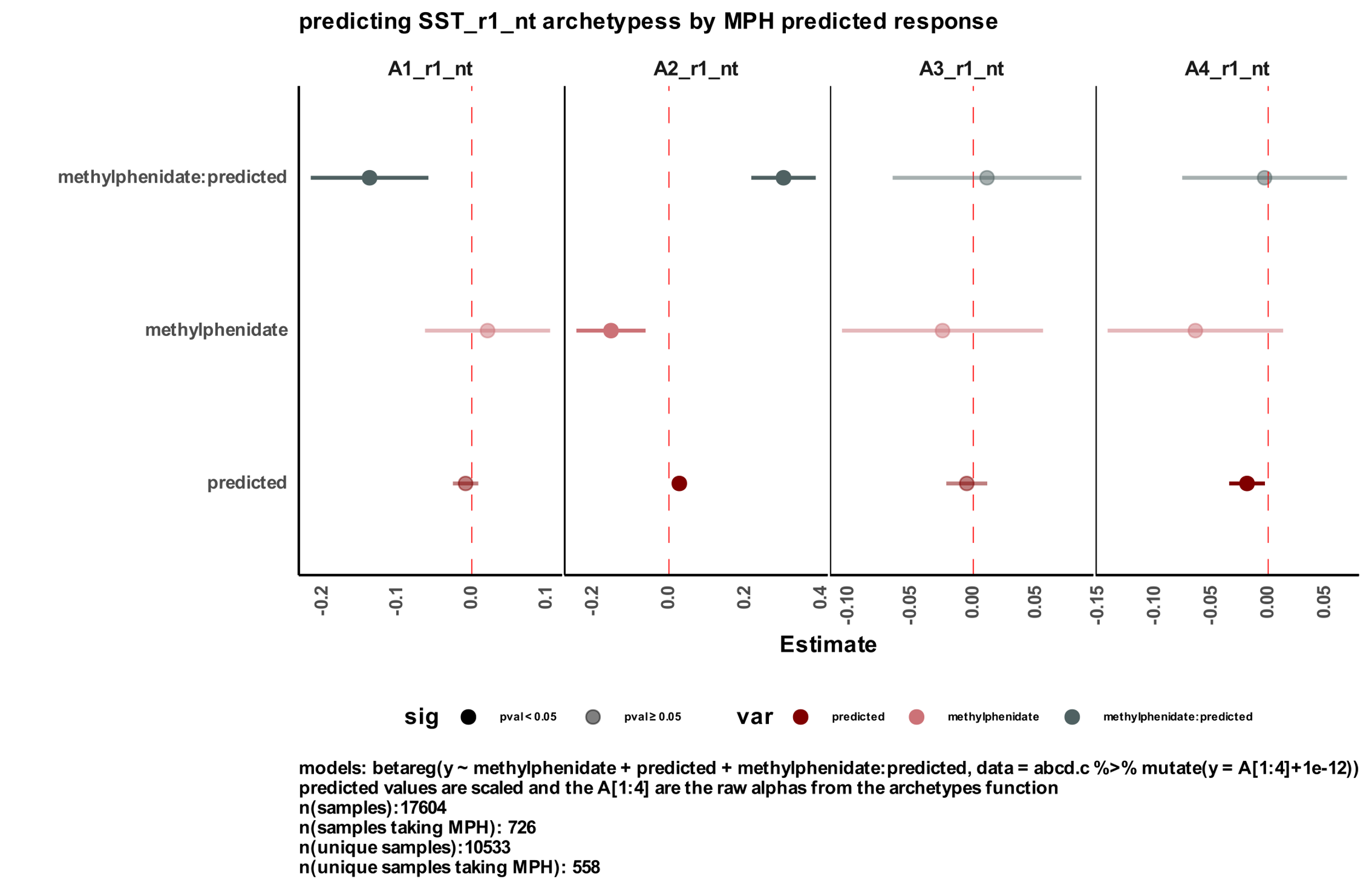
For archetype 2 (A2\_r1\_nt), which was associated with better task performance, participants taking methylphenidate exhibited lower performance. Those predicted to be responders to methylphenidate showed improved task performance, and participants who were both predicted responders and took methylphenidate on the task day performed particularly well. In contrast, the interaction term had a significant negative effect on the first archetype (A1\_r1\_nt), which was negatively correlated with participants' performance in NIH Toolbox tasks.

### 3.6.2. Performance in the NIH Toolbox

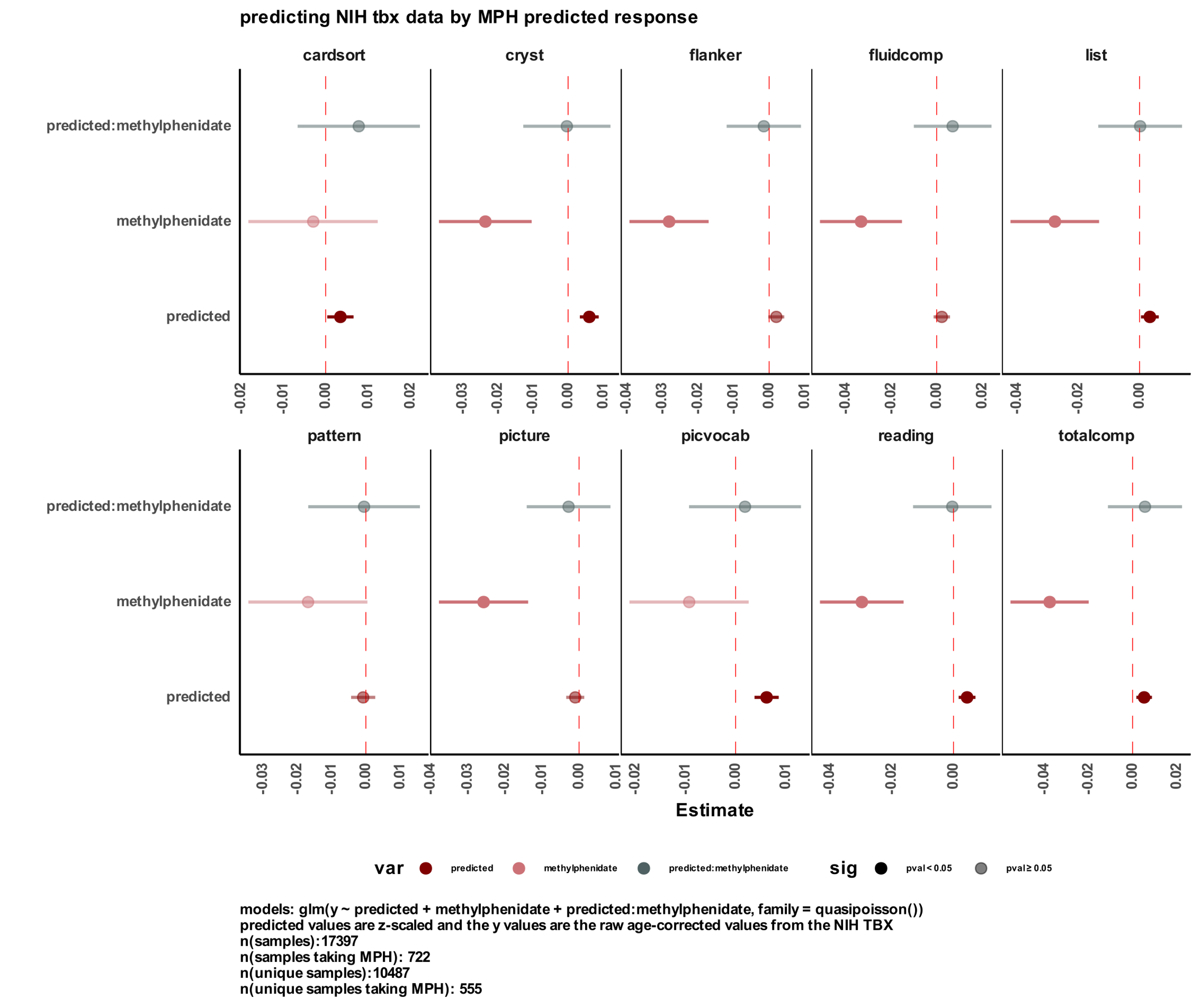
We further investigated the relationship between predicted methylphenidate response and participants' performance in the NIH Toolbox Figure 4-B. Employing a similar regression model, we found significant positive correlations between our predicted responders to methylphenidate and various NIH Toolbox measures, such as cardsort, cryst, list, picvocab, reading, and totalcomp. Conversely, taking methylphenidate was negatively correlated with measures including cryst, flanker, fluidcomp, list, picture, reading, and totalcomp.

These findings highlight the predictive power of our pipeline for drug response and its influence on performance measures in both the SST and NIH Toolbox tasks.

**Figure 4 correlation between predicted methylphenidate response and objective measures of participants' performance on different tasks**  
A) shows the regression model to predict the main four archetypes of participants’ performance in SST by using predicted response to methylphenidate from our pipeline. The regression model was following a beta regression with the following formula: Y ~ methylphenidate + predicted\_response + methylphenidate:predicted\_resposne. B) shows a similar regression model with varying the model to be glm with specifying the family as quasipoisson() and the Y variable to be participants’ performance in different tasks in the NIH Toolbox.



A



B

## 3.7. Method validation in UK Biobank

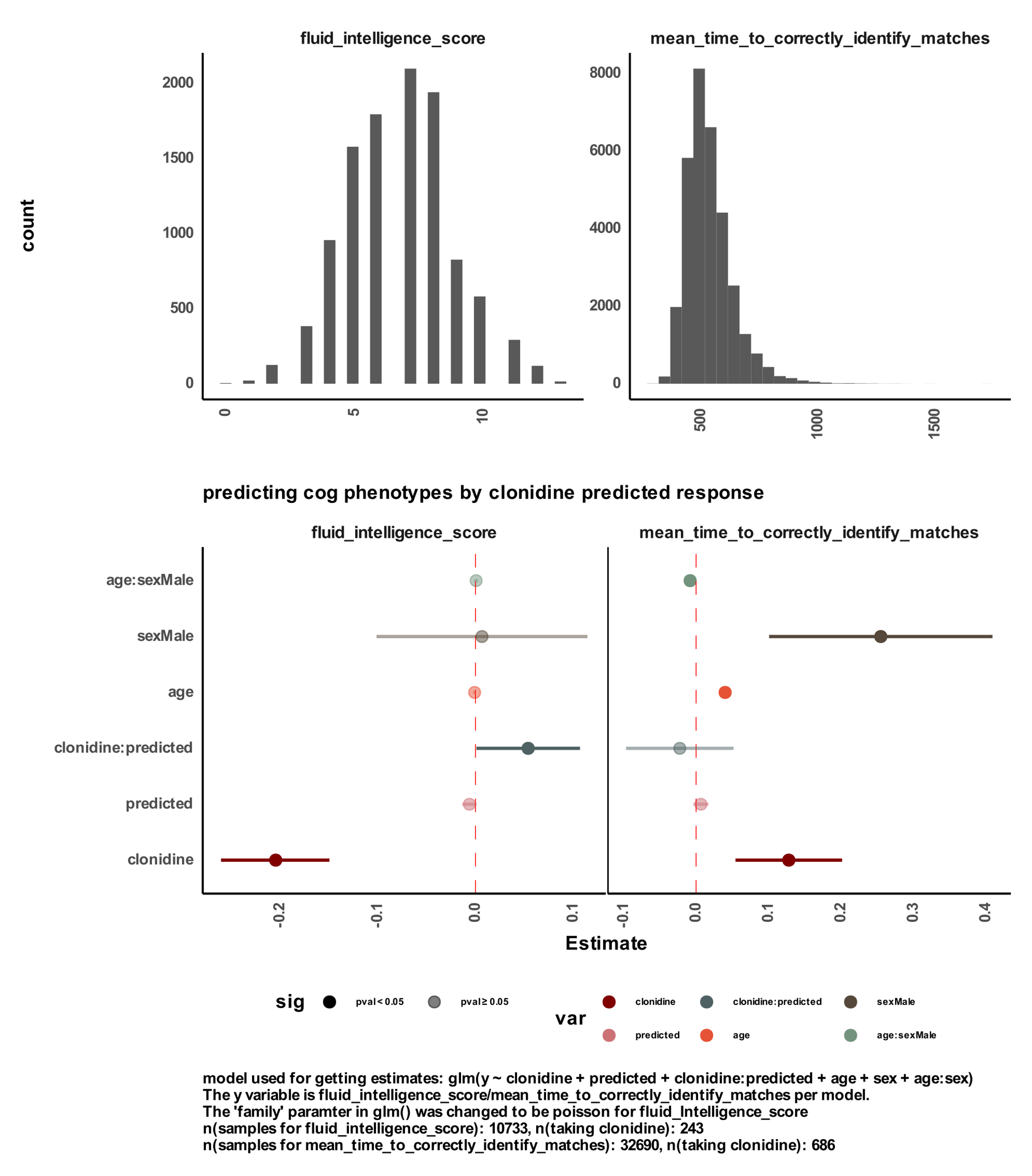
To validate the effectiveness of our drug response prediction pipeline, we conducted a validation study using an independent sample from the UK Biobank. In this study, we focused on predicting participants' response to clonidine, a non-stimulant medication commonly used in the treatment of ADHD. The predicted response to clonidine was then used to predict participants' performance on a fluid intelligence cognitive task and their mean time to correctly identify matches in a [NAME] task Figure 5 (see methods).

We employed a regression model to predict participants' score/time for each task, incorporating several variables including clonidine usage, predicted response to clonidine, age, sex, and their interactions. The formula used for both models was:

Analyzing the results, we observed a significant (nominal p-value: 5.213e-13) negative correlation between clonidine usage and participants' total score on the fluid intelligence task. This suggests that individuals who reported taking clonidine had lower scores on the task, indicating a potential negative impact of clonidine on fluid intelligence. Furthermore, we found a significant (nominal p-value: 0.046) positive correlation for the interaction term between the predicted response to clonidine and the usage of clonidine on the day of task performance. This implies that participants who were predicted to respond positively to clonidine and reported taking it on the same day had better task performance, as reflected by higher scores.

In contrast, we observed a significant (nominal p-value: 6.508e-04) positive correlation between clonidine usage and the mean time taken to correctly identify matches in the [NAME] task. This indicates that participants who reported taking clonidine required more time to perform the task accurately. Although we did not find any significant associations for the predicted response to clonidine and the interaction term in this task, it is noteworthy that the interaction term showed a negative estimate, suggesting that participants who took clonidine and were predicted to respond to it took less time to correctly identify matches.

These findings provide additional support for the reliability and utility of our drug response prediction pipeline, as demonstrated in an independent sample from the UK Biobank. The observed correlations between clonidine usage, predicted response, and task performance underscore the potential impact of clonidine on cognitive functioning. Further research is warranted to explore the underlying mechanisms and optimize treatment strategies for individuals taking clonidine for ADHD or related conditions.



**Figure 5 correlation between predicted clonidine response and objective measures of participants' performance on different tasks.**  
The figure shows a regression model results to predict the participants’ performance in two tasks by using predicted response to clonidine from our pipeline. The regression model was fit by using the formula: Y ~clonidine + predicted + clonidine:predicted + age + sex + age:sex.

# 4. Discussion

The present study aimed to investigate the factors contributing to inter-individual variability in medication response, specifically focusing on the effectiveness of methylphenidate. Leveraging the SPARK cohort, we explored potential communication and cognitive factors that might influence the response to methylphenidate. Our analysis revealed a significant positive correlation between methylphenidate effectiveness and child verbal communication abilities, suggesting that children with better verbal communication skills exhibited a more favorable response to the medication. Additionally, we observed a negative correlation between methylphenidate effectiveness and cognitive impairment, which was further validated through the genetic signal of symbol digit matching polygenic scores (PGS). These findings underscore the role of communication abilities and cognitive functioning in predicting the response to methylphenidate.

To further explore the genetic contributions to medication response, we extended our analysis to an independent study, the ABCD cohort. In this cohort, we investigated the association between polygenic scores for cognitive functions and task performance as a main effect. Interestingly, we observed a significant positive correlation between the cognitive polygenic scores and better performance across various tasks, suggesting that individuals with higher polygenic scores for cognitive functions exhibited enhanced cognitive performance. However, when we examined the interaction between the status of taking methylphenidate at the time of task performance and the cognitive polygenic scores, we encountered unexpected results. The interaction term did not consistently demonstrate the anticipated relationship with performance. Rather, we observed significant interactions in directions that were contrary to our expectations. This discrepancy raises important questions about the predictive value of polygenic scores in capturing the intricate relationship between genetic factors, medication response, and task performance. Our findings highlight the need for further exploration and refinement of genetic markers to better understand the complex interplay between genetics, medication response, and cognitive outcomes.

We then aimed to develop a method for predicting drug response based on genetic profiles and apply it to methylphenidate in the context of neurodevelopmental disorders. By leveraging eQTL weights and imputing transcriptomic profiles for excitatory neurons, we successfully established a pipeline to estimate drug response. Our approach relied on comparing the imputed transcriptome of an individual to the methylphenidate signature derived from the CMap database. This novel methodology allowed us to move beyond subjective reporting and objective measures that require waiting periods for observable drug effects.

The correlations between the imputed transcriptome and drug signature served as indicators of drug effectiveness. Negative correlations suggested that the drug could neutralize the effects of the imputed transcriptome, potentially leading to positive changes in gene expression. Conversely, positive correlations indicated that the drug might exacerbate gene expression in a similar direction to the imputed transcriptome, potentially hampering its therapeutic impact. These findings support the notion that imputed transcriptomic profiles can provide valuable insights into drug response, offering a potential avenue for personalized medicine approaches.

Our pipeline's efficacy was demonstrated by predicting methylphenidate response in participants from the ABCD cohort. We found significant correlations between the predicted response and performance in the SST, which served as an objective measure of attention. Archetype analysis provided a deeper understanding of the relationships between performance and the four archetypes derived from the SST. Notably, higher alpha values in the second archetype (A2\_r1\_nt) indicated better task performance, while higher alpha values in the first, third, and fourth archetypes (A1\_r1\_nt, A3\_r1\_nt, and A4\_r1\_nt) were associated with worse performance. These findings corroborated the relevance of the archetypes and their link to performance measures from the NIH Toolbox.

By fitting regression models to the SST archetypes and including the predicted response to methylphenidate, methylphenidate intake during the task, and their interaction, we further elucidated the relationships between drug response and performance. For archetype 2 (A2\_r1\_nt), which reflected better performance, our results demonstrated a significant positive main effect of most cognitive polygenic scores (PGS) on task performance. This suggests that individuals with higher cognitive abilities, as indicated by the PGS, exhibited better performance in the SST.

Interestingly, taking methylphenidate was associated with lower performance in the task, indicating that the drug might have a negative impact on attention during the task. However, participants predicted to be responders based on our pipeline and who took methylphenidate on the task day demonstrated improved performance. This finding suggests that methylphenidate can enhance task performance in individuals who are likely to respond positively to the drug. These results underscore the potential benefits of using our predictive pipeline to identify individuals who would benefit most from methylphenidate treatment.

Our study also extended its investigation to participants' performance in the NIH Toolbox, using a similar regression model to predict performance by including methylphenidate intake, predicted response to methylphenidate, and their interaction. Consistent with our previous findings, predicted responders to methylphenidate exhibited a significant positive correlation with various NIH Toolbox measures, indicating better performance in tasks such as cardsort, cryst, list, picvocab, reading, and totalcomp. Conversely, taking methylphenidate was negatively correlated with measures including cryst, flanker, fluidcomp, list, picture, reading, and totalcomp.

To validate the efficacy of our predictive pipeline for drug response, we conducted an analysis on a third independent sample derived from the UK Biobank. While the sample size for participants taking methylphenidate was limited, we identified a substantial number of individuals (N: 686) who reported taking clonidine. Notably, clonidine is another medication used for ADHD, but its mechanism of action differs from that of methylphenidate. To ensure accurate predictions, we adjusted our pipeline accordingly by utilizing eQTL weights derived from inhibitory neurons specifically for predicting clonidine response, while maintaining the use of excitatory neuron weights for methylphenidate prediction. Applying our pipeline to the clonidine sample, we employed a regression model that accounted for age, sex, and their interaction, in addition to the main effects and interaction terms associated with clonidine response. Notably, we discovered a significant correlation between the interaction term, which represented the predicted response to clonidine and the status of taking the medication, and participants' total score on the fluid intelligence task. This finding suggests that participants who were predicted to respond to clonidine and reported taking it on the same day of task performance exhibited higher scores on the fluid intelligence task, even after accounting for potential confounding variables.

The results of this study provide valuable insights into the prediction of drug response and its impact on performance measures. By leveraging genetic profiles and imputed transcriptomic data, we successfully predicted methylphenidate response and demonstrated its associations with task performance in the SST and NIH Toolbox tasks. We were also able to demonstrate a significant association between predicted response to clonidine and improved performance on the fluid intelligence task.

# 5. Methods

## 5.1. Genotypes and polygenic scores for SPARK and ABCD

The genotyping and polygenic score methodologies employed in this study were previously documented in our group's publication (Taylor R Thomas, 2023).

### 5.1.1. Genotype quality control

Genotype quality control, imputation, and the computation of genetic principal components and polygenic scores were described in detail in (Taylor R. Thomas, 2022). The ABCD genotypes underwent quality control measures to address missingness and contamination before their official release. Hence, no additional quality control steps were conducted prior to genotype imputation. As for SPARK, we utilized the genotypes from the integrated whole-exome sequencing (iWES1) 2022 Release, along with the SPARK whole-genome sequencing (WGS) Releases 2, 3, and 4. The SPARK iWES1 dataset (N = 69,592) had already undergone quality control, including the removal of samples exhibiting heterozygosity or high missingness. Thus, no further quality control was performed by our team before genotype imputation. Moreover, the SPARK iWES1 dataset provided genetic ancestry assignments based on the 1000 Genomes populations (Genomes Project et al., 2015).

In contrast, SPARK WGS Releases 2 (N = 2,365), 3 (N = 2,871), and 4 (N = 3,684) were not subjected to quality control upon release. Consequently, we conducted quality control using PLINK (Purcell et al., 2007) before proceeding with genotype imputation. Our quality control process involved the exclusion of participants already present in the iWES1 dataset. Additionally, we eliminated variants and participants exhibiting missingness greater than 0.1 and 0.2, respectively. Subsequently, we merged the three WGS releases and excluded any participant whose heterozygosity (F statistic) deviated more than 3 standard deviations from the mean heterozygosity across the three releases. We employed the TopMed reference panel (Taliun et al., 2021) to identify strand flips. Ultimately, the final sample size for WGS 2-4 comprised N = 8,152 individuals.

### 5.1.2. Genotype imputation and merging

We employed the Michigan Imputation Server (Das et al., 2016)to impute the genotypes of ABCD, SPARK iWES1, and quality-controlled SPARK WGS 2-4 datasets to the TopMed reference panel (Taliun et al., 2021). Imputation was carried out using phasing and quality control steps, yielding variants with an imputation quality r2 > 0.3. Following imputation, the variants were filtered to retain only the HapMap SNPs (N = 1,054,330 variants) with an imputation quality r2 > 0.8, utilizing bcftools (Danecek et al., 2021). The variants were then converted from the hg38 to hg19 reference genome using the VCF-liftover tool (<https://github.com/hmgu-itg/VCF-liftover>), and their alleles were normalized accordingly. Finally, the files were merged, and only variants with 0% missingness were retained, resulting in a final dataset size of N = 914,328 variants.

### 5.1.3. Genetic ancestry and principal components

To compute the genetic principal components (PCs), we utilized the bigsnpr package (Prive et al., 2018) following the recommended guidelines outlined by the authors (Prive et al., 2020). For a detailed tutorial, we followed the steps provided in their documentation: [link: <https://privefl.github.io/bigsnpr/articles/bedpca.html>]. In summary, the following steps were performed:

1. We employed the snp\_plinkKINGQC function to identify and exclude related participants using the KING threshold of 2-3.5

2. Principal component analysis (PCA) was conducted using the bed\_autoSVD function solely on unrelated participants.

3. Outliers among the principal components were detected and subsequently removed.

4. The principal components were recalculated.

5. The bed\_projectSelfPCA function was used to project the principal components onto the entire cohort.

To assign genetic ancestry, we performed k-means clustering utilizing the top 40 principal components, with K set to 5 to correspond to the five populations identified in the 1000 Genomes project (Genomes Project et al., 2015). The genetic ancestry labels from the iWES1 dataset were employed to assign labels to the five genetic population clusters.

### 5.1.4. Polygenic score calculations

Polygenic scores (PGS) were calculated using LDpred2 (Prive et al., 2021) and the bigsnpr tools (Prive et al., 2018) in R (R-Core-Team). As SPARK is a family-based dataset, we utilized an external LD reference based on 362,320 individuals from the UK Biobank, provided by the authors of LDpred2. This reference was used to calculate the genetic correlation matrix, estimate heritability, and calculate infinitesimal beta weights.

~~The PGS were derived from the following individual genome-wide association studies (GWAS): ADHD (Demontis et al., 2019) and educational attainment (EA) (Lee et al., 2018). Additionally, the PGS for the five cognitive traits (general cognitive performance, executive functioning, non-verbal reasoning, working memory, and reaction time) were obtained from the same study (de la Fuente et al., 2021).~~

The PGS were derived from the following individual genome-wide association studies (GWAS): ADHD (Demontis et al., 2019) and educational attainment (EA) (Lee et al., 2018). Additionally, the PGS for the eight cognitive traits (fluid intelligence (verbal\_numerical reasoning), matrix pattern completion (matrix), tower rearranging (tower\_rearranging), numeric memory (mem), reaction time (reaction\_time), symbol digit substitution (symbol\_digit), trail making (trail\_making), and a *G* factor score (gFactor)) were obtained from the same study.

To account for potential confounding effects, we conducted linear regression residualization by incorporating the first five genetic principal components. Subsequently, the PGS were Z-scaled to have a mean of 0 and a standard deviation (σ) of 1.

## 5.2. SPARK methylphenidate effectiveness correlations

### 5.2.1. SPARK participants

## For our analysis of methylphenidate (MPH) effectiveness, we utilized a survey that was administered to a subset of SPARK participants. The survey, completed by parents, focused on aspects such as sleep patterns, eating habits, and gastrointestinal issues in their children. Among the questions, one specifically targeted the medical history of the child and assessed the effectiveness of methylphenidate as a binary variable: 0 for not effective and 1 for effective. The classification was solely based on the perceived effectiveness of the drug without considering any potential adverse effects. Participants were asked about the drug's effectiveness, irrespective of whether they were currently taking it or had taken it in the past.

We utilized the data collected from this question as the predicted response for drug effectiveness and compared it to other surveys conducted by SPARK for the same participants. The sample included 1460 participants, with a mean age of 160.580 months (SD = 65.164).

### 5.2.2. social communication questionnaire

In addition to the effectiveness survey, parents also reported data from the Social Communication Questionnaire (SCQ) for the same participants. The SCQ is a reliable screening tool used to assess social communication abilities and identify potential indicators of autism spectrum disorder (ASD). Comprising 40 structured questions, the SCQ evaluates various aspects of social interaction, communication skills, and behavioral patterns commonly associated with ASD. It covers areas such as eye contact, gestures, facial expressions, reciprocal conversation, interests, and repetitive behaviors. The SCQ offers a comprehensive evaluation of social communication abilities, making it valuable for research investigations and clinical assessments of ASD. Each question in the SCQ is answered with a yes or no response.

To examine the relationship between MPH effectiveness and individual SCQ questions, we conducted Fisher's Exact tests using the 'fisher.test()' function in R for each question independently. We extracted the odds ratio and nominal p-value from these tests. The table below summarizes the sample included in the analysis and reports the questions with significant odds ratios associated with MPH effectiveness.

**Table 1 social communication questionnaire associations with MPH effectiveness**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Question | No | Yes | Odds Ratio | p-value |
| MPH effect | 435 | 683 | - | - |
| q01\_phrases | 72 | 1046 | 2.1870605 | 0.001637545 |
| q02\_conversation | 285 | 833 | 1.3964096 | 0.016828588 |
| q04\_inappropriate\_question | 266 | 852 | 1.3452012 | 0.043571607 |
| q15\_odd\_ways | 355 | 763 | 0.6515883 | 0.001553941 |
| q17\_injured\_deliberately | 501 | 617 | 0.7372904 | 0.016111320 |
| q24\_nod\_head | 366 | 752 | 1.3243570 | 0.031188530 |
| q25\_shake\_head | 321 | 797 | 1.4925137 | 0.002880866 |

### 5.2.3. cognitive impairment

To explore the association between MPH effectiveness and cognitive impairment among the participants, we conducted a Fisher's Exact test. The cognitive impairment variable was derived as a binary vector from the SPARK cohort: 0 denoting no cognitive impairment and 1 indicating cognitive impairment. This variable served as a predictor for multiple measurements in the participants. The table below presents a summary of the sample included in this section.

**Table 2 cognitive impairment counts for participants with reported MPH effectiveness**

|  |  |  |
| --- | --- | --- |
| Question | No | Yes |
| MPH effect | 548 | 862 |
| derived\_cog\_impair | 593 | 817 |

### 5.2.4. polygenic scores

After adjusting the polygenic scores (PGS) for all ABCD and SPARK participants using the first five genetic principal components, we specifically included participants who had reported MPH effectiveness in the SPARK dataset. Subsequently, we examined the correlation between PGS and MPH effectiveness using the 'cor()' function in R. The reported values in Figure 1-C represent the Spearman's rank-based correlation coefficients along with their nominal p-values.

## 5.3. ABCD cohort

The present study utilized data from the Adolescent Brain Cognitive Development (ABCD) dataset, a comprehensive and longitudinal study conducted in the United States. The primary objective of the ABCD dataset is to examine brain development in relation to behavioral, cognitive, and social outcomes among adolescents. The dataset comprises a diverse sample of over 11,000 children, initially aged 9 to 10 years at baseline, who are followed through adolescence with multiple assessments conducted over several years. These assessments cover a wide range of domains, including neuroimaging, neurocognitive testing, psychosocial questionnaires, and physical health measures. By incorporating these various assessments, the ABCD study provides a multidimensional dataset that facilitates investigations into different aspects of brain development and associated outcomes during the crucial period of adolescence.

### 5.3.1. Stop signal fMRI task.

In the ABCD dataset, the Stop Signal fMRI Task (SST) was utilized to assess inhibitory control and impulsivity in adolescents. This task measures the ability to inhibit prepotent responses by presenting participants with a series of go trials where they need to respond quickly to a specific stimulus. On certain trials, a stop signal is presented after the go signal, indicating that participants should inhibit their response. By manipulating the timing of the stop signal, researchers can estimate the stop signal reaction time (SSRT), which reflects the individual's ability to successfully inhibit their responses. The SST provides valuable insights into the neural mechanisms underlying inhibitory control and impulsivity, contributing to the understanding of cognitive processes related to behavioral regulation in the ABCD study.

During the SST, typically two trial runs are conducted to familiarize participants with the task and obtain reliable measurements of inhibitory control. The first trial run serves as an introduction to the task instructions and structure, allowing participants to become comfortable and understand the requirements for response inhibition. For this analysis, we focused on the data from the first run since most participants did not complete the second run.

We extracted specific variables of interest from the evaluated SST data as measures of attention. These variables include the number of correct and incorrect go answers within the specified time, the number of correct and incorrect go answers given late, the number of correct and incorrect stop answers, the number of no responses on go trials, and the number of stop signal doesn't stop trials.

These variables were found to have a significant correlation with the interview age of the participants. To account for age, we conducted a regression analysis using the glm() function in R, fitting the formula glm(y ~ interview\_age, family = quasipoisson()) and obtaining the residuals.

To reduce the dimensionality of the data, we utilized the stepArchetypes() function from the archetypes package in R. By examining the scree plot, we selected the first four archetypes for inclusion. The alphas obtained from the output represent values bounded between 0 and 1 for each participant. Higher values for archetype 2 (A2\_r1\_nt) indicate better performance in the task, while higher values in archetypes 1 (A1\_r1\_nt), 3 (A3\_r1\_nt), and 4 (A4\_r1\_nt) indicate poorer performance in the task.

### 5.3.2. NIH Toolbox data

To further understand the meaning of the archetypes derived from the SST in the ABCD dataset, we examined their correlation with the National Institutes of Health (NIH) Toolbox data collected for the same participants. The ABCD dataset incorporates a range of cognitive assessments from the NIH Toolbox, which are widely used in research and clinical settings. Within the ABCD dataset, several NIH Toolbox measures were employed, including the Pattern Comparison Processing Speed Test, List Sorting Working Memory Test, Flanker Inhibitory Control and Attention Test, Picture Sequence Memory Test (Crest), and Dimensional Change Card Sort Test. These measures cover various cognitive domains, including visual processing speed, working memory, inhibitory control, attention, episodic memory, and cognitive flexibility. For our analysis, we assessed the Spearman's correlation between the four main archetypes derived from the SST and the age-corrected NIH Toolbox data. We included the NIH Toolbox data after applying an age correction. This correlation analysis provides insights into the relationship between the archetypes obtained from the SST and the cognitive processes measured by the NIH Toolbox.

### 5.3.3. Polygenic scores

After adjusting the polygenic scores (PGS) for all participants in the SPARK and ABCD cohorts using the first five genetic principal components, we focused on participants from the ABCD dataset who had reported performance data in the first run of the SST. We then examined the correlation between the PGS, and the archetypes derived from the SST using a regression model to predict the alpha values of each archetype using the PGS as a main effect. The model also included other variables as shown in thew model formula below:

The Y variable in the formula represents the alphas of each archetype. The methylphenidate variable in the model indicates whether participants were taking methylphenidate during the fMRI task. The last variable in the model represents the interaction term between the PGS and methylphenidate status. To fit the regression model, we used the betareg() function, taking care to add a small value to the alpha (Y variable) of the archetypes. This adjustment was made to ensure the model did not violate the assumption of a beta distribution with values in the (0,1) interval. The added small value was chosen arbitrarily, such as 1e-12. By including this small value, we aimed to retain the beta distribution shape while keeping the values within the desired interval.

## 5.4. Transcriptome imputation

To investigate how different genetic variants influence gene expression in specific cell types, we utilized cis-eQTL weights obtained from (Bryois et al., 2022). These weights were assigned to each variant-gene pair and were derived using a specific statistical model that accounted for various factors such as genotype, cell type, principal components, study, disease status, and individual-specific random effects. The eQTLs were derived by the following model as defined by the authors:

The study provided cis-eQTL weights for eight different cell types, including astrocytes, endothelial cells, excitatory neurons, inhibitory neurons, microglia, oligodendrocytes, OPCs (oligodendrocyte progenitor cells), and pericytes. However, for our investigation into methylphenidate response, we focused solely on excitatory neurons. To narrow down the list of eQTLs, we filtered them based on a false discovery rate (FDR) threshold of less than 0.05, resulting in a final set of 328,061 unique significant eQTLs out of the original 2,438,542 eQTLs.

To impute gene expression in excitatory neurons for individuals, we performed a simple matrix multiplication using the following formula:

The cell\_type\_Weights\_Matrix represents the weights matrix derived from (Bryois et al., 2022). It has dimensions MxN, where M represents the number of variants and N represents the number of genes. The Genotypes\_Matrix, on the other hand, has rows representing individuals and columns representing genotypes. The values in this matrix indicate the type of genotype alleles found for each individual (0 for homozygous reference, 1 for heterozygous, and 2 for homozygous alternate).

For our analysis, we used the Genotypes\_Matrix specific to participants from the Adolescent Brain Cognitive Development (ABCD) study. The ABCD study is one of the largest long-term studies funded by the National Institutes of Health (NIH) that focuses on brain development and child health in the United States. It includes approximately 11,000 children aged 9-10 years at the beginning of the study.

After imputing transcriptome data for the ABCD participants, we further standardized the imputed values on a gene level by z-scaling them, ensuring a mean of 0 and a standard deviation (σ) of 1. This normalization step allows for easier interpretation and comparison of gene expression levels across individuals.

## 5.5. Drug response

### 5.5.1. calculating polygenic drug response

To assess individuals' response to methylphenidate (MPH), we calculated the correlation between their imputed transcriptome in excitatory neurons and the drug signature obtained from the connectivity map (CMap)[[3]](#footnote-3) database developed by the BROAD Institute. The CMap database provides a measure of cellular response to different perturbagens (including drugs) in various cell lines.

The drug signature matrix for each drug is a 1xG matrix, where G represents the total number of genes with measured gene expression after introducing the drug to the cell line. The gene expression values in the signature matrix range from 0 to 1, representing fluorescence intensity. For our analysis, we used the fitted data from the CMap, which controls for the cell line and drug dose, ensuring that the signature reflects the drug effect specifically.

To calculate the correlation, we employed the cor() function in R, specifically using Spearman rank-based correlation. This correlation measures how different an individual's imputed gene expression is compared to the drug signature. Before calculating the correlation, we accounted for genes that had a measured change in the drug signature but did not have imputed transcriptome values (either due to a lack of eQTLs found in the sample or the gene not having significant eQTLs with weights). In these cases, we added those genes to the imputed transcriptome vector with a value of 0.

The calculated correlations serve as a measure of how an individual's imputed gene expression differs from the drug signature. Based on these correlations, individuals were categorized as positive responders to the drug if their imputed tissue expression had a negative correlation with the drug perturbation signature. Conversely, individuals were categorized as negative responders if they had a positive correlation. In this context, a negative correlation indicates that the drug is expected to reverse the imputed transcriptome, thereby normalizing the gene expression.

To ensure consistency in the interpretation, we multiplied the calculated correlations by -1 to reverse the sign. Thus, a positive value in the correlation indicates that an individual is predicted to be a responder to methylphenidate (MPH).

### 5.5.2. comparing predicted polygenic MPH response to participants’ performance

After calculating the polygenic methylphenidate (MPH) response, we proceeded to examine its correlation with participants' performance in two different tasks: the Stop Signal Task (SST) and the NIH Toolbox measures. We employed regression models to assess the relationship, using the following formula:

In the case of the SST, we considered each of the four main archetypes independently as the Y variable in the formula mentioned above. The methylphenidate variable represents a binary variable indicating whether the participant reported taking methylphenidate on the day they performed the fMRI task. To fit the regression model, we followed the same outline in predicting archetypes of the SST using PGS as the main effect (mentioned in the polygenic scores section of ABCD). We used the betareg() function, taking care to add a small value to the alpha (Y variable) of the archetypes. This adjustment was made to ensure the model did not violate the assumption of a beta distribution with values in the (0,1) interval. The added small value was chosen arbitrarily, such as 1e-12. By including this small value, we aimed to retain the beta distribution shape while keeping the values within the desired interval.

For the regression model predicting the NIH Toolbox measures, we encountered non-normal distribution in the variables. Therefore, we utilized the following parameters in the glm() function to fit the model:

glm(Y ~ methylphenidate + predicted\_MPH\_response + methylphenidate:predicted\_MPH\_response, family = quasipoisson())

Similar to the SST analysis, we modified the Y variable to correspond to each specific NIH Toolbox measure.

By fitting these regression models, we obtained estimates for each variable, allowing us to examine the relationship between methylphenidate, predicted MPH response, and participants' performance in the SST and NIH Toolbox measures.

5.6. Validation in a sample from UK Biobank

5.6.1. sample participants

In this study, we utilized data from the UK Biobank, a large-scale population-based cohort study that has made significant contributions to biomedical research. The UK Biobank dataset encompasses a wealth of information on a diverse range of health-related measures, including genetic data, imaging data, and self-reported health information. It consists of over 500,000 participants aged 40-69 years recruited from across the United Kingdom. This rich and extensive dataset enables investigations into various aspects of health and disease, providing an invaluable resource for conducting large-scale genetic and phenotypic analyses. By leveraging the UK Biobank data, we were able to conduct additional analyses and validate our findings, thereby enhancing the robustness and generalizability of our study results.

Initially, we searched for participants who reported taking methylphenidate but found a limited number available for inclusion in our analysis. However, we identified a substantial sample size of participants reporting the use of clonidine (N: 703) among other ADHD medications. To assess participants' cognitive performance, we focused on two specific tasks: fluid intelligence and identifying matches. While the UK Biobank offers a range of cognitive tasks, we chose these two due to the availability of data for a large number of participants taking clonidine (N: 243 for the fluid intelligence task and N: 686 for the identifying matches task, recording mean time). Therefore, our analysis included participants who completed these two tasks, resulting in a combined sample of 32,786 individuals with diverse ethnic backgrounds.

5.6.2. genotypes

After selecting a subset of participants for analysis, we filtered the genotype data from the UK Biobank to include only the samples relevant to our study. We utilized the imputed genotypes v.3 from the UK Biobank without any additional imputation for missing genotyped variants, similar to our approach in the SPARK and ABCD cohorts. We then imputed missing genotype data on a per-participant basis, replacing them with the median value for each variant.

5.6.3. transcriptome imputation

To impute transcriptome data, we followed the same steps as those performed in the ABCD cohort, with one modification. Instead of using eQTL weights derived from excitatory neurons, we utilized eQTL weights from inhibitory neurons. This adjustment was made based on the distinct mechanism of action between methylphenidate and clonidine. Given that methylphenidate is a stimulant, we selected excitatory neurons for the transcriptome imputation. Conversely, as clonidine is not a stimulant and is believed to exert inhibitory effects, we focused on studying its impact on inhibitory neurons. We applied a similar pipeline to that used in the ABCD cohort, with the only change being the source of the weights.

5.6.4. drug response

To evaluate drug response, we obtained the clonidine transcriptomic drug signature from the Connectivity Map (CMap) and computed Spearman rank-based correlation coefficients between the imputed transcriptome profiles of individual participants and the clonidine signature. For a more detailed explanation, please refer to the drug response section in the ABCD cohort.

5.6.5. comparing predicted polygenic clonidine response to participants’ performance

After calculating the polygenic clonidine response, we examined its correlation with participants' performance in two distinct tasks: fluid intelligence and matching identification. To assess the relationship, we employed regression models using the following formula:

The "clonidine" variable represented a binary indication of whether participants reported taking clonidine during the verbal interview. The "predicted\_clonidine\_response" variable represented a z-scaled measure of the predicted response derived from our pipeline. Age, sex, and their interaction were included in the model as they exhibited correlations with participants' cognitive task performance. The "Y" variable in the model was adjusted to represent either the fluid intelligence score or the mean time taken to correctly identify matches.

For fitting the regression model to both tasks, we utilized the glm() function in R and adjusted the "family" parameter to match the data distribution. Specifically, we used the Poisson family for the fluid intelligence score and left the default family parameter value for the mean time taken to correctly identify matches, which was z-scaled.

# 6. Limitations

Several limitations should be considered when interpreting the findings of this study. Firstly, the use of subjective reporting for MPH effectiveness and social communication abilities relies on the accuracy of parent reports, which can be influenced by various factors such as recall bias or subjective perceptions. Objective measures, although employed in the validation stage, still have inherent limitations, and may not fully capture the complexity of drug response in real-world settings.

Another limitation is the reliance on transcriptomic imputation for predicting drug response. While this approach provides valuable insights into gene expression patterns, it relies on the assumption that imputed values accurately represent the true transcriptomic profile of the individual. It is also worth noting that the eQTL weights were derived from postmortem samples. The imputation process itself introduces inherent uncertainty and potential biases, which may affect the accuracy of the predicted drug response.

Furthermore, the study's sample size, although representative of the targeted populations, may limit the generalizability of the findings. The SPARK and ABCD cohorts, while providing valuable data, may not fully capture the diversity and heterogeneity present in the broader population of individuals with neurodevelopmental disorders.

Additionally, the regression models used in this study are based on associations and correlations and do not establish causality. Other factors beyond the genetic and transcriptomic profiles, such as environmental influences or individual-specific physiological processes, could contribute to the observed drug response variations.

Lastly, it is worth noting that the study focused specifically on methylphenidate response in individuals with neurodevelopmental disorders. The generalizability of the findings to other medications or different patient populations requires further investigation.

These limitations highlight the need for caution when interpreting the results and suggest avenues for future research. Larger and more diverse cohorts, rigorous validation of predictive models, and longitudinal studies assessing the long-term effects of predicted drug response are essential to further refine and validate the proposed method for personalized pharmacological management in neurodevelopmental disorders.

# 7. Acknowledgements

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive DevelopmentSM (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9-10 and follow them over 10 years into early adulthood. The ABCD Study® is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at <https://abcdstudy.org/federal-partners.html>. A listing of participating sites and a complete listing of the study investigators can be found at<https://abcdstudy.org/consortium_members/>. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators.

The ABCD data repository grows and changes over time. The ABCD data used in this report came from [NIMH Data Archive Digital Object Identifier 10.15154/1523041]. DOIs can be found at https://nda.nih.gov/abcd.

The SPARK data can be obtained at SFARI Base: <https://base.sfari.org>.

# 8. Data availability

The data used in this study is subject to data access and sharing policies set forth by the respective cohorts and databases. The SPARK (Simons Foundation Powering Autism Research for Knowledge) cohort data can be accessed through an application process via the SPARK website (https://sparkforautism.org/). Researchers interested in accessing the ABCD (Adolescent Brain Cognitive Development) cohort data can submit data access requests to the ABCD Study Data Repository (https://nda.nih.gov/abcd).

The imputation weights used in this study are derived from publicly available resources. The specific sources and versions of the genetic data are detailed in the Methods section. Researchers can access these resources directly to obtain the necessary genetic information.

The connectivity map (CMap) drug perturbation signature data used to compute drug response correlations can be obtained from the BROAD Institute's CMap website (https://clue.io/).

Due to the sensitive nature of the participant data, including clinical and behavioral information, access to individual-level data from the SPARK and ABCD cohorts may require adherence to ethical and privacy considerations. Researchers interested in accessing the individual-level data should consult the respective cohort data access policies and procedures, ensuring compliance with relevant legal and ethical guidelines.

In summary, access to the data used in this study is subject to the policies and procedures of the SPARK and ABCD cohorts, as well as the availability of publicly accessible resources. Researchers can refer to the cohort websites and associated repositories for detailed information on data access and availability.

**9. Code availability**

To promote transparency and reproducibility, the code used for data analysis, statistical modeling, and generation of results and figures can be found here [link].

# 10. Funding

[ask Jake]

# 11. 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.

# 12. Author contributions

ME contributed to the study design, generating the imputed transcriptome, predicting drug response, processing the data, performing the statistical analyses, and writing the manuscript. TRT contributed to processing the data, genotype imputations, generating the polygenic scores, and writing the manuscript. LCG contributed to processing the data. JJM contributed to study design and writing the manuscript.

# 13. References

American-Academy-of-Pediatrics. (2022). Common ADHD Medications & Treatments for Children. Retrieved 06/19/2023, from <https://www.healthychildren.org/English/health-issues/conditions/adhd/Pages/Determining-ADHD-Medication-Treatments.aspx>

Belle, D. J., & Singh, H. (2008). Genetic factors in drug metabolism. *Am Fam Physician*, *77*(11), 1553-1560. <https://www.ncbi.nlm.nih.gov/pubmed/18581835>

Bryois, J., Calini, D., Macnair, W., Foo, L., Urich, E., Ortmann, W., Iglesias, V. A., Selvaraj, S., Nutma, E., Marzin, M., Amor, S., Williams, A., Castelo-Branco, G., Menon, V., De Jager, P., & Malhotra, D. (2022). Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders. *Nat Neurosci*, *25*(8), 1104-1112. <https://doi.org/10.1038/s41593-022-01128-z>

CDC, C. f. D. C. a. P. (2022, August 9, 2022). *Data and Statistics About ADHD*. CDC. Retrieved June 16th from <https://www.cdc.gov/ncbddd/adhd/data.html>

Chang, J. C., Lin, H. Y., Lv, J., Tseng, W. I., & Gau, S. S. (2021). Regional brain volume predicts response to methylphenidate treatment in individuals with ADHD. *BMC Psychiatry*, *21*(1), 26. <https://doi.org/10.1186/s12888-021-03040-5>

Cortese, S., Adamo, N., Del Giovane, C., Mohr-Jensen, C., Hayes, A. J., Carucci, S., Atkinson, L. Z., Tessari, L., Banaschewski, T., Coghill, D., Hollis, C., Simonoff, E., Zuddas, A., Barbui, C., Purgato, M., Steinhausen, H. C., Shokraneh, F., Xia, J., & Cipriani, A. (2018). Comparative efficacy and tolerability of medications for attention-deficit hyperactivity disorder in children, adolescents, and adults: a systematic review and network meta-analysis. *Lancet Psychiatry*, *5*(9), 727-738. <https://doi.org/10.1016/S2215-0366(18)30269-4>

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *Gigascience*, *10*(2). <https://doi.org/10.1093/gigascience/giab008>

Das, S., Forer, L., Schonherr, S., Sidore, C., Locke, A. E., Kwong, A., Vrieze, S. I., Chew, E. Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P. R., Iacono, W. G., Swaroop, A., Scott, L. J., Cucca, F., Kronenberg, F., Boehnke, M., . . . Fuchsberger, C. (2016). Next-generation genotype imputation service and methods. *Nat Genet*, *48*(10), 1284-1287. <https://doi.org/10.1038/ng.3656>

de Faria, J. C. M., Duarte, L. J. R., Ferreira, L. A., da Silveira, V. T., Menezes de Padua, C., & Perini, E. (2022). "Real-world" effectiveness of methylphenidate in improving the academic achievement of Attention-Deficit Hyperactivity Disorder diagnosed students-A systematic review. *J Clin Pharm Ther*, *47*(1), 6-23. <https://doi.org/10.1111/jcpt.13486>

de la Fuente, J., Davies, G., Grotzinger, A. D., Tucker-Drob, E. M., & Deary, I. J. (2021). A general dimension of genetic sharing across diverse cognitive traits inferred from molecular data. *Nat Hum Behav*, *5*(1), 49-58. <https://doi.org/10.1038/s41562-020-00936-2>

Demontis, D., Walters, R. K., Martin, J., Mattheisen, M., Als, T. D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Baekvad-Hansen, M., Cerrato, F., Chambert, K., Churchhouse, C., Dumont, A., Eriksson, N., Gandal, M., Goldstein, J. I., Grasby, K. L., Grove, J., . . . Neale, B. M. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet*, *51*(1), 63-75. <https://doi.org/10.1038/s41588-018-0269-7>

Dunnenberger, H. M., Crews, K. R., Hoffman, J. M., Caudle, K. E., Broeckel, U., Howard, S. C., Hunkler, R. J., Klein, T. E., Evans, W. E., & Relling, M. V. (2015). Preemptive clinical pharmacogenetics implementation: current programs in five US medical centers. *Annu Rev Pharmacol Toxicol*, *55*, 89-106. <https://doi.org/10.1146/annurev-pharmtox-010814-124835>

Evans, W. E., & Johnson, J. A. (2001). Pharmacogenomics: the inherited basis for interindividual differences in drug response. *Annu Rev Genomics Hum Genet*, *2*, 9-39. <https://doi.org/10.1146/annurev.genom.2.1.9>

Genomes Project, C., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., & Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68-74. <https://doi.org/10.1038/nature15393>

Grazioli, S., Rosi, E., Mauri, M., Crippa, A., Tizzoni, F., Tarabelloni, A., Villa, F. M., Chiapasco, F., Reimers, M., Gatti, E., Bertella, S., Molteni, M., & Nobile, M. (2021). Patterns of Response to Methylphenidate Administration in Children with ADHD: A Personalized Medicine Approach through Clustering Analysis. *Children (Basel)*, *8*(11). <https://doi.org/10.3390/children8111008>

Ingelman-Sundberg, M., Mkrtchian, S., Zhou, Y., & Lauschke, V. M. (2018). Integrating rare genetic variants into pharmacogenetic drug response predictions. *Hum Genomics*, *12*(1), 26. <https://doi.org/10.1186/s40246-018-0157-3>

Kimko, H. C., Cross, J. T., & Abernethy, D. R. (1999). Pharmacokinetics and clinical effectiveness of methylphenidate. *Clin Pharmacokinet*, *37*(6), 457-470. <https://doi.org/10.2165/00003088-199937060-00002>

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T. A., Bowers, P., Sidorenko, J., Karlsson Linner, R., Fontana, M. A., Kundu, T., Lee, C., Li, H., Li, R., Royer, R., Timshel, P. N., Walters, R. K., Willoughby, E. A., . . . Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet*, *50*(8), 1112-1121. <https://doi.org/10.1038/s41588-018-0147-3>

Pliszka, S. R. (1998). Comorbidity of attention-deficit/hyperactivity disorder with psychiatric disorder: an overview. *J Clin Psychiatry*, *59 Suppl 7*, 50-58. <https://www.ncbi.nlm.nih.gov/pubmed/9680053>

Prive, F., Arbel, J., & Vilhjalmsson, B. J. (2021). LDpred2: better, faster, stronger. *Bioinformatics*, *36*(22-23), 5424-5431. <https://doi.org/10.1093/bioinformatics/btaa1029>

Prive, F., Aschard, H., Ziyatdinov, A., & Blum, M. G. B. (2018). Efficient analysis of large-scale genome-wide data with two R packages: bigstatsr and bigsnpr. *Bioinformatics*, *34*(16), 2781-2787. <https://doi.org/10.1093/bioinformatics/bty185>

Prive, F., Luu, K., Blum, M. G. B., McGrath, J. J., & Vilhjalmsson, B. J. (2020). Efficient toolkit implementing best practices for principal component analysis of population genetic data. *Bioinformatics*, *36*(16), 4449-4457. <https://doi.org/10.1093/bioinformatics/btaa520>

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, *81*(3), 559-575. <https://doi.org/10.1086/519795>

R-Core-Team. R: A Language and Environment for Statistical Computing. <https://www.R-project.org/>

Relling, M. V., & Evans, W. E. (2015). Pharmacogenomics in the clinic. *Nature*, *526*(7573), 343-350. <https://doi.org/10.1038/nature15817>

Roden, D. M., Wilke, R. A., Kroemer, H. K., & Stein, C. M. (2011). Pharmacogenomics: the genetics of variable drug responses. *Circulation*, *123*(15), 1661-1670. <https://doi.org/10.1161/CIRCULATIONAHA.109.914820>

Scharfe, C. P. I., Tremmel, R., Schwab, M., Kohlbacher, O., & Marks, D. S. (2017). Genetic variation in human drug-related genes. *Genome Med*, *9*(1), 117. <https://doi.org/10.1186/s13073-017-0502-5>

Storebo, O. J., Pedersen, N., Ramstad, E., Kielsholm, M. L., Nielsen, S. S., Krogh, H. B., Moreira-Maia, C. R., Magnusson, F. L., Holmskov, M., Gerner, T., Skoog, M., Rosendal, S., Groth, C., Gillies, D., Buch Rasmussen, K., Gauci, D., Zwi, M., Kirubakaran, R., Hakonsen, S. J., . . . Gluud, C. (2018). Methylphenidate for attention deficit hyperactivity disorder (ADHD) in children and adolescents - assessment of adverse events in non-randomised studies. *Cochrane Database Syst Rev*, *5*(5), CD012069. <https://doi.org/10.1002/14651858.CD012069.pub2>

Taliun, D., Harris, D. N., Kessler, M. D., Carlson, J., Szpiech, Z. A., Torres, R., Taliun, S. A. G., Corvelo, A., Gogarten, S. M., Kang, H. M., Pitsillides, A. N., LeFaive, J., Lee, S. B., Tian, X., Browning, B. L., Das, S., Emde, A. K., Clarke, W. E., Loesch, D. P., . . . Abecasis, G. R. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*, *590*(7845), 290-299. <https://doi.org/10.1038/s41586-021-03205-y>

Taylor R Thomas, L. G. C., Jacob J Michaelson. (2023). Environmental adversity moderates polygenic score effects on childhood behavioral problems in the United States [Preprint]. *medRxiv*. <https://doi.org/10.1101/2023.06.16.23291504>

Taylor R. Thomas, A. J. T., Amy M. Pearlman, Katherine L. Imborek, Ji Seung Yang, John F. Strang, Jacob J. Michaelson. (2022). Polygenic scores clarify the relationship between mental health and gender diversity. *medRxiv*. <https://doi.org/10.1101/2021.11.22.21266696>

Vertessen, K., Luman, M., Swanson, J. M., Bottelier, M., Stoffelsen, R., Bet, P., Wisse, A., Twisk, J. W. R., & Oosterlaan, J. (2023). Methylphenidate dose-response in children with ADHD: evidence from a double-blind, randomized placebo-controlled titration trial. *Eur Child Adolesc Psychiatry*. <https://doi.org/10.1007/s00787-023-02176-x>

Weinshilboum, R. (2003). Inheritance and drug response. *N Engl J Med*, *348*(6), 529-537. <https://doi.org/10.1056/NEJMra020021>

Wilke, R. A., & Dolan, M. E. (2011). Genetics and variable drug response. *JAMA*, *306*(3), 306-307. <https://doi.org/10.1001/jama.2011.998>

# 14. Supplementary tables

These are extra table (quick & dirty). I’m not sure if I will include all of these or not. Also, the formatting might be changed.

**Table S 1 demographics for ABCD participants that performed the stop signal fMRI task.**

|  |  |  |
| --- | --- | --- |
| count | Mean of interview age in months (SD) | Sex |
| 19262 | 128.693 (14.099) | M & F |
| 10157 | 129.084 (14.19) | M |
| 9105 | 128.257 (13.985) | F |

**Table S 2 mean count for participants' performance in the stop signal task**

|  |  |  |
| --- | --- | --- |
| Question label | Question details | Mean count of answers (SD) |
| tfmri\_sst\_r1\_beh\_crgo\_nt | Number of correct GO answers in time. | 126.388 (22.015) |
| tfmri\_sst\_r1\_beh\_crlg\_nt | Number of correct GO answers and late. | 3.580 (4.249) |
| tfmri\_sst\_r1\_beh\_incrgo\_nt | Number of incorrect GO answers in time. | 11.812 (12.051) |
| tfmri\_sst\_r1\_beh\_incrlg\_nt | Number of incorrect GO answers and late. | 0.302 (0.997) |
| tfmri\_sst\_r1\_beh\_nrgo\_nt | Number of NO RESPONSE on GO trials. | 7.919 (16.326) |
| tfmri\_sst\_r1\_beh\_crs\_nt | Number of correct STOP answers. | 16.052 (3.186) |
| tfmri\_sst\_r1\_beh\_incrs\_nt | Number of incorrect STOP answers. | 13.167 (3.687) |
| tfmri\_sst\_r1\_beh\_ssds\_nt | Number of STOP SIGNAL DOESN’T STOP trials. | 0.782 (1.218) |

**Table S 3 demographics of SPARK participants with reported methylphenidate effectiveness**

|  |  |  |
| --- | --- | --- |
| count | Mean of interview age in months (SD) | Sex |
| 1460 | 160.580 (65.164) | M & F |

**Table S 4 counts for methylphenidate effectiveness in SPARK participants**

|  |  |
| --- | --- |
| MPH effect = 0 | MPH effect = 1 |
| 574 | 886 |

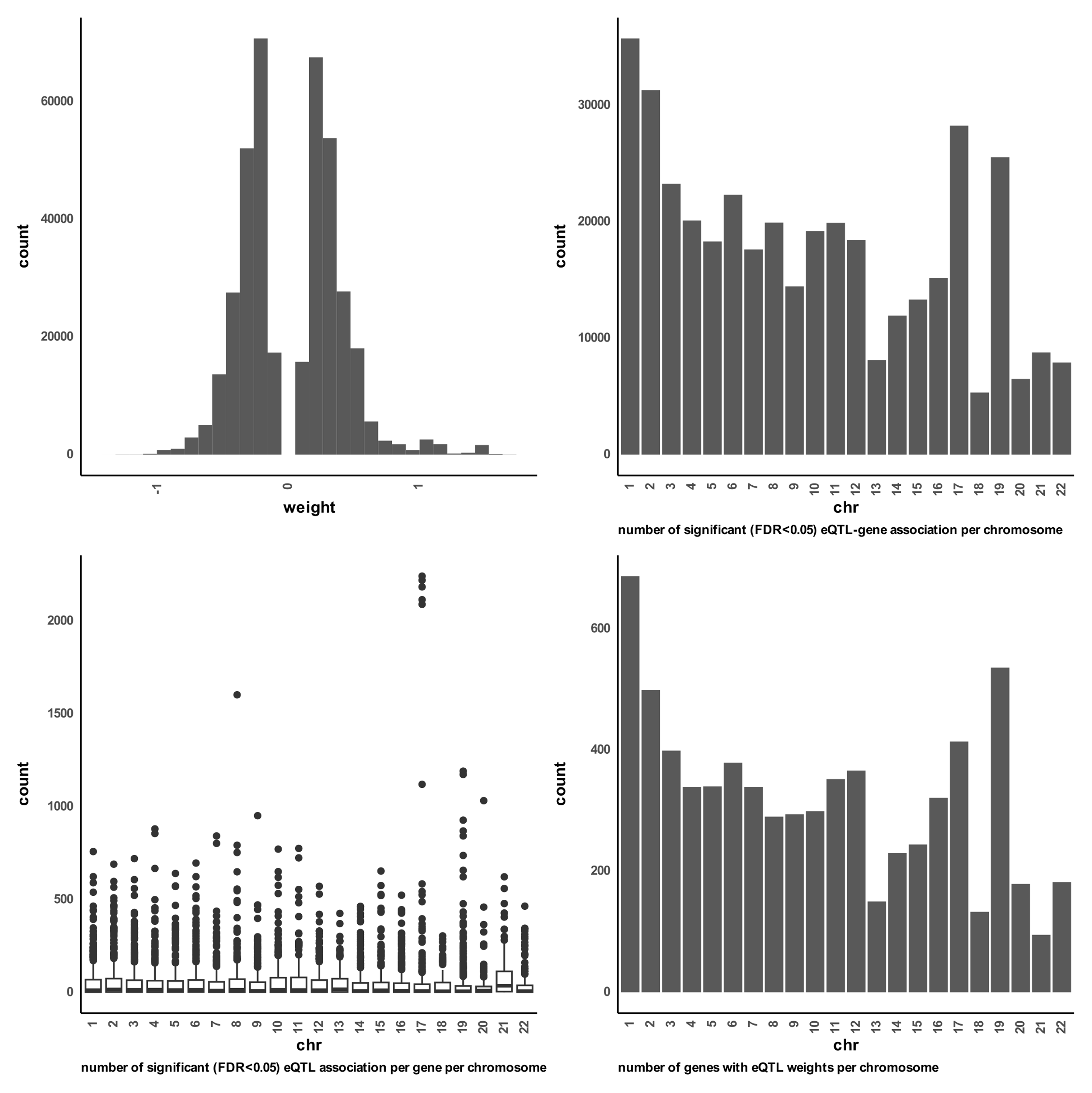
15. supplementary figs

**A**

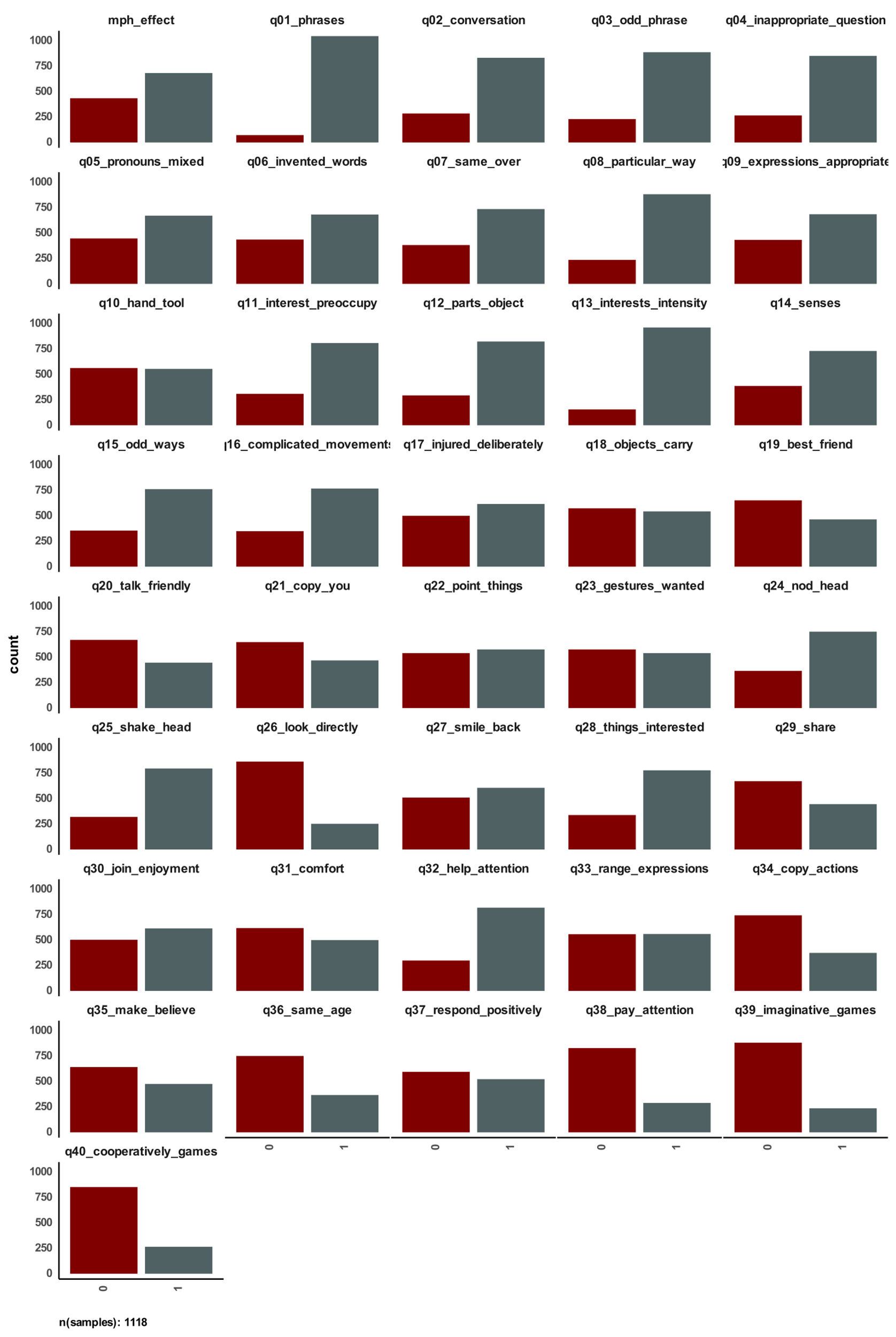
**C**

**D**

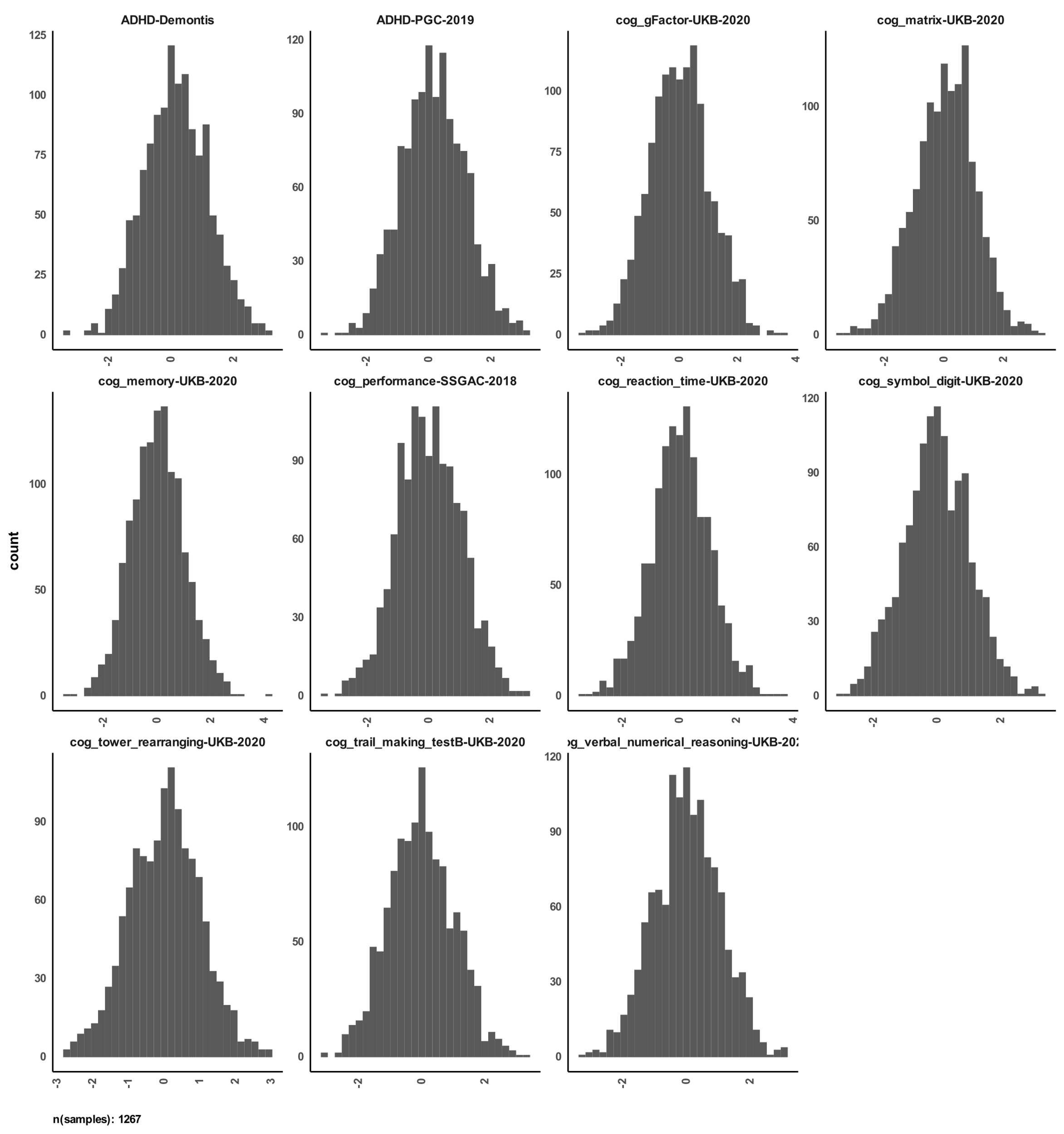
**B**



**Figure S 1 distributions of excitatory eQTL weights.**   
A) a histogram distribution of all significant (FDR < 0.05) gene-eQTL weights of all chromosomes. B) boxplot distribution of number of eQTL associations found per gene on each chromosome. C) boxplot for distribution of number of significant (FDR < 0.05) gene-eQTL associations per chromosome. D) distribution of number of unique genes found with significant (FDR < 0.05) gene-eQTL associations per chromosome.



**Figure S 2 distribution of answers to social communication questionnaire in SPARK and used MPH effectiveness label.**   
The figure shows number of YES (=1) and NO (=0) answers per question in the social communication questionnaire for all samples with reported methylphenidate effectiveness.



**Figure S 3 distribution of calculated polygenic scores for SPARK participants with reported methylphenidate effectiveness.**



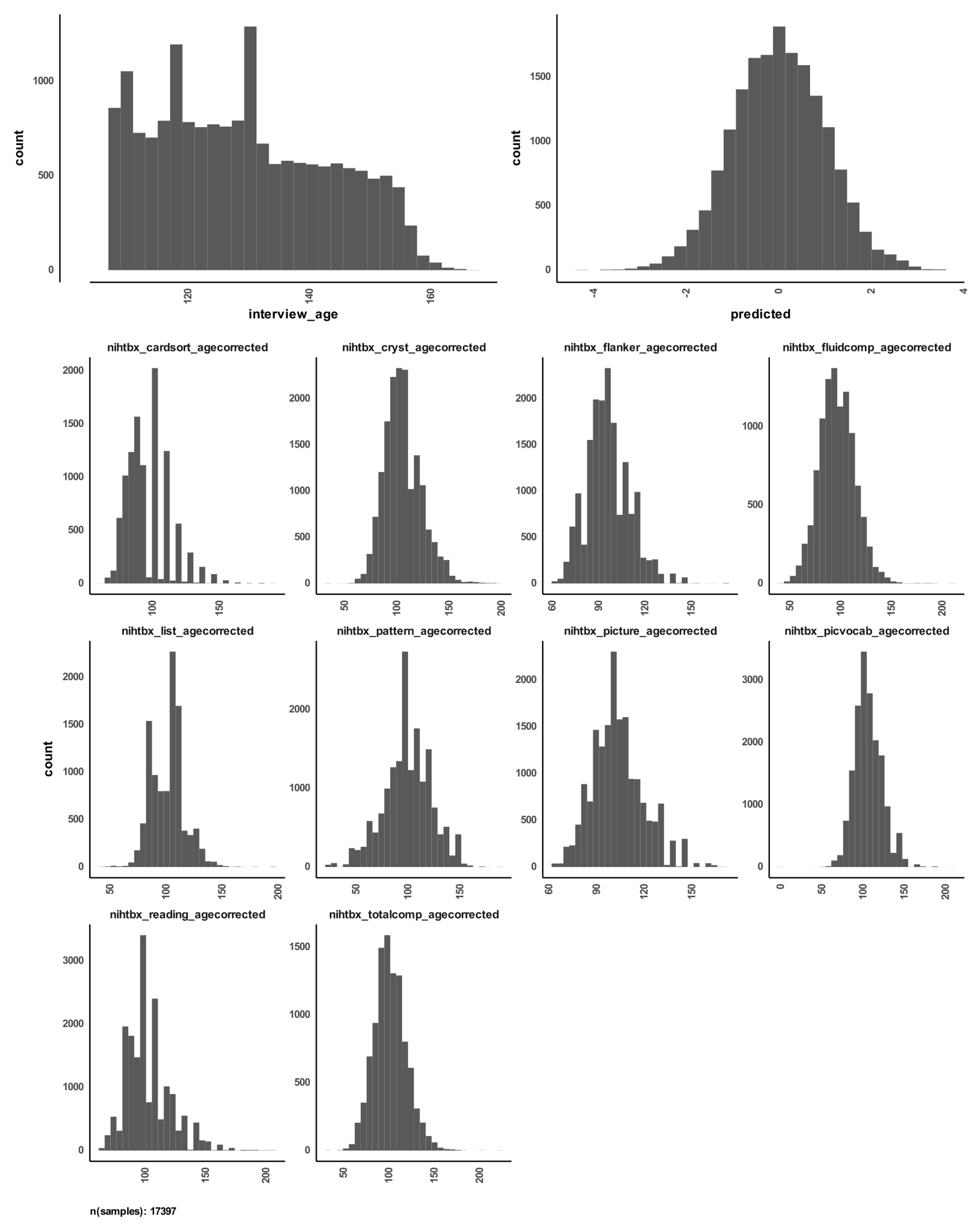
A

B

C

D

**Figure S 4 distributions of archetypes, age, predicted MPH response, and polygenic scores for ABCD participants performing SST.**  
A) distribution of archetype alphas for the main four archetypes of SST. B) distribution of participants’ age that performed the SST. C) distribution of predicted methylphenidate response from our pipeline. D) distribution of calculated polygenic scores for participants.

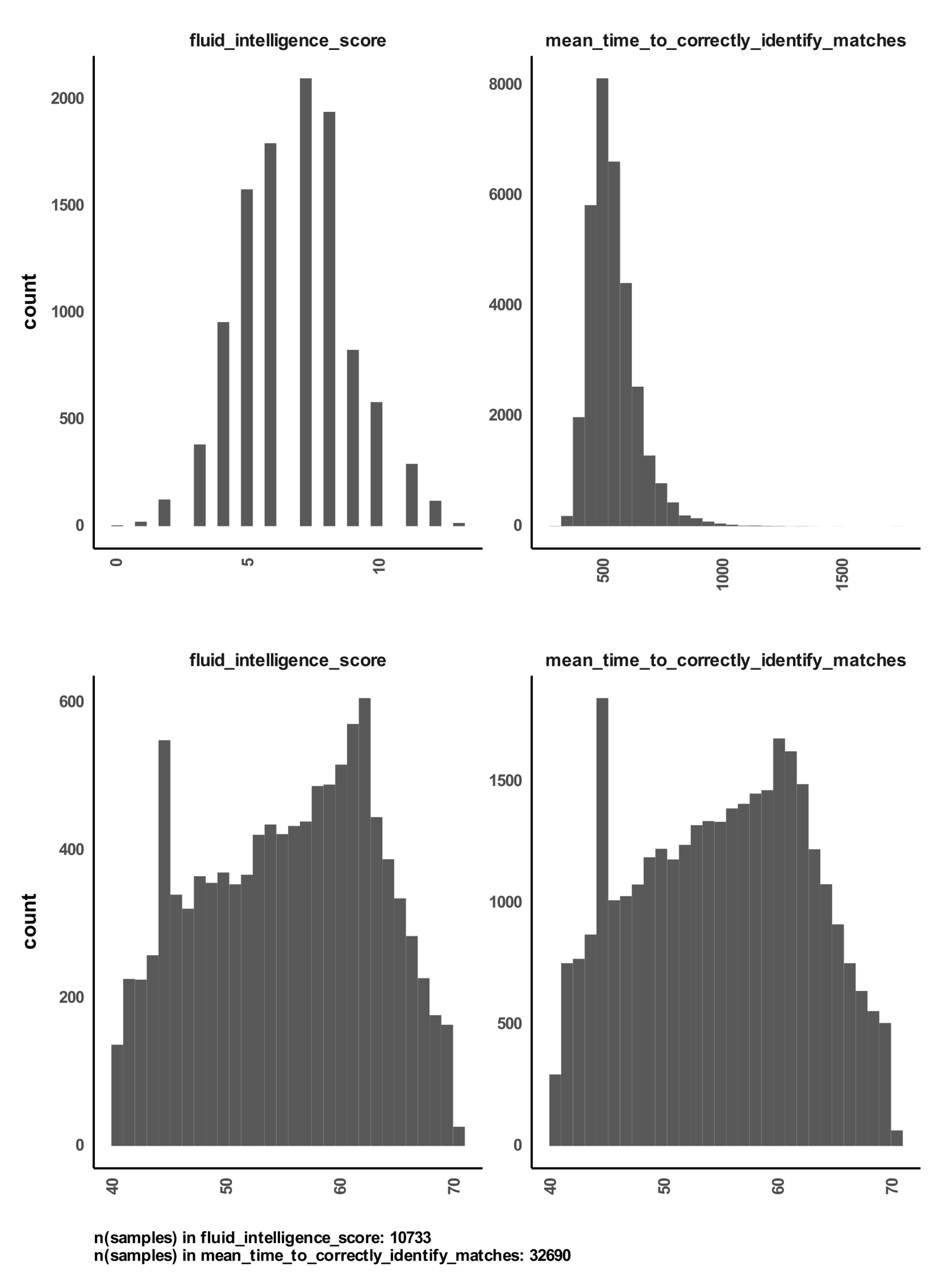


A

B

C

**Figure S 5 distributions of age, predicted MPH response, and total scores per task for ABCD participants performing the NIH Toolbox tasks.**   
A) distribution of participants’ age in months. B) distribution of predicted methylphenidate response from our pipeline. C) distribution of participants’ total score per task in the NIH Toolbox.



A

B

**Figure S 6 distributions of total score and mean time for cognitive tasks from UK Biobank**A) shows the score distribution in the fluid intelligence task, and the mean time to correctly identify matches in another task. B) shows the age distribution of samples used per task.

1. https://sparkforautism.org [↑](#footnote-ref-1)
2. https://abcdstudy.org [↑](#footnote-ref-2)
3. <https://www.broadinstitute.org/connectivity-map-cmap> [↑](#footnote-ref-3)