Evidence for the Placenta-Brain Axis: Multi-Omic Kernel Aggregation Predicts Intellectual and Social Impairment in Children Born Extremely Preterm

Hudson P Santos Jr., PhD[1,2*#], Arjun Bhattacharya, PhD[3*], Robert M Joseph, PhD[4], Lisa Smeester, PhD[2,5,6], Karl CK Kuban, MD[7], Carmen J Marsit, PhD[8], T. Michael O'Shea, MD[9], Rebecca C Fry, PhD[2,5,6]

[1] Biobehavioral Laboratory, School of Nursing, University of North Carolina, Chapel Hill, NC, USA

[2] Institute for Environmental Health Solutions, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA

[3] Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California-Los Angeles, Los Angeles, CA, USA

[4] Department of Anatomy and Neurobiology, Boston University School of Medicine, Boston, MA, USA

[5] Curriculum in Toxicology and Environmental Medicine, University of North Carolina, Chapel Hill, NC, USA

[6] Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA

[7] Department of Pediatrics, Division of Pediatric Neurology, Boston University Medical Center, Boston, MA, USA

[8] Department of Environmental Health, Emory University, Atlanta, GA 30322, USA

[9] Department of Pediatrics, School of Medicine, University of North Carolina, Chapel Hill, NC, USA.

[*]Co-first authorship

[#]Corresponding author: Hudson P Santos Jr.

544 Carrington Hall, Campus Box 7460, Chapel Hill, NC 27599-7460. Email: hsantosj@email.unc.edu

Abstract

Background: Children born extremely preterm are at heightened risk for intellectual and social impairment, including Autism Spectrum Disorder (ASD). There is increasing evidence for a key role of the placenta in prenatal developmental programming, suggesting that the placenta may explain origins of neurodevelopmental outcomes.

Methods: We examined associations between placental genomic and epigenomic profiles and assessed their ability to predict intellectual and social impairment at age 10 years in 379 children from the Extremely Low Gestational Age Newborn (ELGAN) cohort. Assessment of intellectual ability (IQ) and social function was completed with the Differential Ability Scales-II (DAS-II) and Social Responsiveness Scale (SRS), respectively. Examining IQ and SRS allows for studying ASD risk beyond the diagnostic criteria, as IQ and SRS are continuous measures strongly correlated with ASD. Genome-wide mRNA, CpG methylation and miRNA were assayed with the Illumina Hiseq 2500, HTG EdgeSeq miRNA Whole Transcriptome Assay, and Illumina EPIC/850K array, respectively. We conducted genome-wide differential mRNA/miRNA and epigenome-wide placenta analyses. These molecular features were integrated for a predictive analysis of IQ and SRS outcomes using kernel aggregation regression. We lastly examined associations between ASD and the genomically-predicted component of IQ and SRS.

Results: Genes with important roles in placenta angiogenesis and neural function were associated with intellectual and social impairment. Kernel aggregations of placental multi-omics strongly predicted intellectual and social function, explaining approximately 8% and 12% of the variance in SRS and IQ scores via cross-validation, respectively. Predicted in-sample SRS and IQ showed significant positive and negative associations with ASD case-control status.

Limitations: The ELGAN is a cohort of children born pre-term, andgeneralization may be affected by unmeasured confounders associated with low gestational age. We conducted external validation of predictive models, though the sample size of the out-sample dataset ($N = 49$) and the scope of the available placental datasets are limited. Further validation of the models is merited.

Conclusions: Aggregating information from biomarkers within and between molecular data types improves prediction of complex traits like social and intellectual ability in children born extremely preterm, suggesting that traits influenced by the placenta-brain axis may be omnigenic.

Keywords: prenatal neurodevelopmental programming, social and cognitive impairment, placental gene regulation, epigenome-wide association, differential expression analysis, multi-omic aggregation

Background

Despite substantial research efforts to elucidate the etiology of neurodevelopmental impairment [1], little is known about genomic and epigenomic factors influencing trajectories of neurodevelopment, such as those associated with preterm delivery [2]. Children born extremely preterm are at increased risk not only for intellectual impairment but also for Autism Spectrum Disorder (ASD) [3,4], often accompanied by intellectual disability. In addition, preterm-born children have consistently been observed to manifest social difficulties (e.g., fewer prosocial behaviors) in childhood and adolecense that do not meet diagnostic criteria for ASD [5].

The placenta is posited as a critical determinant of both immediate and long-lasting neurodevelopmental outcomes in children [1]. The placenta is involved in hormone and neurotransmitter production and transfer of nutrients to the fetus, thus having direct influence on brain development. This connection between the placenta and the brain is termed the placenta-brain axis [6]. Epidemiological and animal studies have linked genomic and epigenomic alterations in the placenta with neurodevelopmental disorders and normal neurobehavioral development [7–9]. For example, the Markers of Autism Risk in Babies: Learning Early Signs (MARBLES) study has identified differentially methylated region containing putative fetal brain enhancer between in placentas from ASD ($N = 24$) and typically developing (n = 23) children [10]. However, identifying genomic signatures of risk for neurodevelopmental disorders such as ASD in placenta is a challenging. Further study of molecular interactions representing the placenta-brain axis may advance our understanding of fetal mechanisms involved in aberrant neurodevelopment [6].

Most prior studies have investigated single molecular levels of the placenta genome or epigenome, precluding analysis of possible interactions that could be linked to neurodevelopmental outcomes. Examining only a single molecular feature, or a single type of features even at a genomic scale can still result in much unexplained variation in phenotype due to potentially important interactions between multiple features [11,12]. This observation is in line with Boyle *et al*.'s omnigenic model [13,14], which proposes that gene regulatory networks are so highly interconnected that a large portion of the heritability of complex traits can be explained by effects on genes outside core pathways. Molecular integration to

identify pathways for fetal neurodevelopment in children has been unexplored but may prove to be insightful in associations with complex diseases [15].

We conducted a genome-wide analysis of DNA methylation, miRNA, and mRNA expression in the placenta, examining individual associations with social and intellectual impairment at 10 years of age in children from the Extremely Low Gestational Age Newborn (ELGAN) study [16]. We then combined the genomic and epigenomic data to identify correlative networks of placental genomic and epigenomic biomarkers predictive of social and intellectual impairment as continuous scales, thus allowing us to study neurodevelopmental difficulties beyond the ASD diagnostic categories [17]. To assess the convergent validity of our behavioral findings, we also examined the association of social and intellectual impairment in relation to ASD diagnoses [18]. To our knowledge, this is the first study to use multiple placental molecular signatures to predict intellectual and social impairment, which may inform a framework for predicting risk of adverse neurocognitive and neurobehavioral outcomes in young children.

Methods

ELGAN recruitment and study participants

From 2002-2004, women who gave birth at under 28 weeks gestation at one of 14 medical centers across five U.S. states enrolled in the ELGAN study [16]. The Institutional Review Board at each participating institution approved study procedures. Included were 411 of 889 children with both placental molecular analysis and a 10-year follow-up assessment.

Social and cognitive function and ASD at 10 years of age

Trained child psychologist examiner [5,19] evaluated general cognitive ability (IQ) with the School-Age Differential Ability Scales-II (DAS-II) Verbal and Nonverbal Reasoning subscales [20]. The Social Responsiveness Scale (SRS) was used to assess severity of ASD-related social deficits in 5 subdomains: social awareness, social cognition, social communication, social motivation, and autistic mannerisms [21]. We used the gender-normed T-score (SRS-T; intended to correct gender differences observed in normative samples) as continuous measure of social deficit [22]. All participants were assessed for ASD

[18]. Diagnostic assessment of ASD was conducted with three well-validated measures, administered sequentially. First, the Social Communication Questionnaire (SCQ) was administered to screen for potential ASD, using a score ≥ 11 to increase sensitivity relative to the standard criterion score of ≥ 15 [18,23]. For children who screened positive on the SCQ criterion, we conducted the Autism Diagnostic Interview–Revised (ADI-R) with the primary caregiver [24]. All children who met ADI-R criteria for ASD, or who had a prior clinical diagnosis of ASD and/or exhibited symptoms of ASD during cognitive testing according to the site psychologist) were then assessed with the Autism Diagnostic Observation Schedule, Second Version (ADOS-2), which served as the criterion measure of ASD in this study [25]. All ADOS-2 administrations were independently scored by a second rater with autism diagnostic and ADOS-2 expertise. In cases of scoring disagreements, consensus was reached via discussion between raters. Item-by-item inter-rater agreement for the 14 ADOS-2 diagnostic algorithm scores was on average 0.93 ($SD = 0.12$). These developmental assessment procedures and all relevant test scores for ASD and intellectual function are reported in a prior publication [19].

Placental DNA and RNA extraction

After delivery, placentas were biopsied under sterile conditions. We collected a piece of the chorion, representing the fetal side of the placenta [26]. More specifically, placentas were placed in a sterilized basin and biopsied by pulling back the amnion to expose the chorion at the midpoint of the longest distance between the cord insertion and edge of the placental disk. A sample from the *fetal side* of the placenta was removed by applying traction to the chorion and underlying trophoblast tissue. The specimen was placed in a cryogenic vial and immersed in liquid nitrogen. To preserve DNA and RNA integrity, specimens were stored at -80^{o}C until processed. For processing, a 0.2g subsection of the placental tissue was cut from the frozen biopsy and washed with sterile 1x phosphate-buffered saline to remove any remaining blood. Samples were homogenized using a lysis buffer, and the homogenate was separated into aliquots. This process was detailed in a prior publication [27]. Nucleic acids were extracted from the homogenate using AllPrep DNA/RNA/miRNA Universal kit (Qiagen, Germany). The quantity and quality of DNA and RNA were analyzed using the NanoDrop 1000 spectrophotometer and its integrity verified by the Agilent 2100 BioAnalyzer.

143

144 *Epigenome-wide placental DNA methylation*

145 Extracted DNA sequences were bisulfate-converted using the EZ DNA methylation kit (Zymo Research,
146 Irvine, CA) and followed by quantification using the Infinium MethylationEPIC BeadChip (Illumina, San
147 Diego, CA), which measures CpG loci at a single nucleotide resolution, as previously described [26–29].
148 Quality control and normalization were performed resulting in 856,832 CpG probes from downstream
149 analysis, with methylation represented as the average methylation level at a single CpG site (β-value)
150 [27,30–32]. DNA methylation data was imported into R for pre-processing using the *minfi* package [30].
151 Quality control was performed at the sample level, excluding samples that failed and technical duplicates;
152 411 samples were retained for subsequent analyses. Functional normalization was performed with a
153 preliminary step of normal-exponential out-of band (*noob*) correction method [33] for background
154 subtraction and dye normalization, followed by the typical functional normalization method with the top
155 two principal components of the control matrix [31,34]. Quality control was performed on individual probes
156 by computing a detection P value and excluded 806 (0.09%) probes with non-significant detection ($P >$
157 0.01) for 5% or more of the samples. A total of 856,832 CpG sites were included in the final analyses.
158 Lastly, the *ComBat* function was used from the *sva* package to adjust for batch effects from sample plate
159 [83]. The data were visualized using density distributions at all processing steps. Each probe measured
160 the average methylation level at a single CpG site. Methylation levels were calculated and expressed as
161 β values (β = intensity of the methylated allele (M))/(intensity of the unmethylated allele (U) + intensity of
162 the methylated allele (M) + 100). β values were logit transformed to M values for statistical analyses [35].

163

164 *Genome-wide placental mRNA and miRNA expression*

165 mRNA expression was determined using the Illumina QuantSeq 3′ mRNA-Seq Library Prep Kit, a method
166 with high strand specificity. mRNA-sequencing libraries were pooled and sequenced (single-end 50 bp)
167 on one lane of the Illumina Hiseq 2500. mRNA were quantified through pseudo-alignment with *Salmon*
168 v.14.0 [36] mapped to the GENCODE Release 31 (GRCh37) reference transcriptome. miRNA expression
169 profiles were assessed using the HTG EdgeSeq miRNA Whole Transcriptome Assay (HTG Molecular
170 Diagnostics, Tucson, AZ). miRNA were aligned to probe sequences and quantified using the HTG

171 EdgeSeq System [37]. Genes and miRNAs with less than 5 counts for each sample were filtered,
172 resulting in 11,224 genes and 2,047 miRNAs for downstream analysis. Distributional differences between
173 lanes were first upper-quartile normalized [38]. Unwanted technical and biological variation (e.g. tissue
174 heterogeneity) was then estimated using *RUVSeq* [39], where we empirically defined transcripts not
175 associated with outcomes of interest as negative control housekeeping probes [40]. One dimension of
176 unwanted variation was removed from the variance-stabilized transformation of the gene expression data
177 using the *limma* package [40–43]
178
179 *Statistical Analysis*
180 All code and functions used in the statistical analysis can be found at https://github.com/bhattacharya-a-
181 bt/multiomics_ELGAN.
182
183 *Correlative analyses between SRS, IQ, and ASD*
184 Associations among SRS scores, IQ and ASD were assessed using Pearson correlations with estimated
185 95% confidence intervals, and the difference in distributions of SRS and IQ across ASD case-control was
186 assessed using Wilcoxon rank-sum tests. Associations between demographic variables (race, sex,
187 maternal age, number of gestational days, maternal smoking status, placental inflammation, birth weight
188 Z-score and mother's insurance) with SRS and IQ were assessed using multivariable regression,
189 assessing the significance of regression parameters using Wald tests of significance and adjusting for
190 multiple testing with the Benjamini-Hochberg procedure [44].
191
192 *Genome-wide molecular associations with SRS and IQ*
193 Once associations between SRS and IQ and ASD were confirmed, we utilized continuous SRS and IQ
194 measures as the main outcomes of interest. Associations between mRNA expression or miRNA
195 expression with SRS and IQ were estimated through a negative binomial linear model using *DESeq2* [43].
196 Epigenome-wide associations (EWAS) of CpG methylation sites with outcomes were assessed using
197 robust linear regression [45] with test statistic modification through an empirical Bayes procedure [42],
198 described previously [27]. Both the differential mRNA and miRNA expression and EWAS models

controlled for the following covariates: race, age, sex, number of gestational age days, birth weight Z-score, and education level of the mother. Multiple testing was adjusted for using the Benjamini-Hochberg procedure [44].

Placental multi-molecular prediction of SRS and IQ

We next assessed how well an aggregate of one or more of the molecular datasets (CpG methylation, mRNA expression, and miRNA expression) predicted continuous SRS and IQ scores. The analytical scheme is summarized in **Figure 1**, using 379 samples with data for all three molecular datasets (DNA methylation, miRNA, and mRNA). Briefly, we first adjusted the outcome variables and molecular datasets for above noted demographic and clinical covariates using *limma* [46] to account for associations between the outcomes and these coviarates in the eventual predictive models. Next, to model the covariance between samples within a single molecular profile, we aggregated the molecular datasets with thousands of biomarkers each into a *molecular kernel* matrix. A *molecular kernel* matrix represents the inter-sample similarities in a given molecular profile (**Supplementary Methods**). A linear or non-linear kernel aggregation may aid in prediction of complex traits by capturing non-additive effects [47–50], which represents a sizable portion of phenotypic variation [51,52]. Using all individual, pairwise, and triplet-wise combinations of molecular kernel matrices, we fitted predictive models of SRS and IQ based on linear mixed modeling [50] or kernel regression least squares (KRLS) [53] and assessed predictive performance with McNemar's adjusted R^2 via Monte Carlo cross validation [54]. We also optimized predictive models for the number of included biomarkers per molecular profile. Extensive model details, as well as alternative models considered, are detailed in **Supplemental Methods.**

Validation in external dataset

Lack of studies that consider placental mRNA, CpG methylation and miRNA data with long-term child neurodevelopment limit the ability to extablish external validation. We obtained one external placental CpG methylation dataset from the Markers of Autism Risk in Babies-Learning Early Signs (MARBLES) cohort [10]. To assess out-of-sample performance of kernel models for methylation, we downloaded MethylC-seq data for 47 placenta samples, 24 of which identified as ASD cases (NCBI Gene Expression

Omnibus accession numbers GSE67615) [10]. β-values for DNA methylation were extracted from BED files and transformed into M-values with an offset of 1 [35], and used the best methylation-only predictive model to predict SRS and IQ in these 47 samples, as detailed in **Supplemental Methods**.

Correlative networks

In the final KRLS predictive models for both IQ and SRS including all three molecular profiles, we extracted the top 50 most predictive (largest point-wise effect sizes) CpGs, miRNAs, and mRNAs of SRS and IQ. A sparse correlative network was inferred among these biomarkers that links biomarkers based on the strength of correlative signals using graphical lasso in *qgraph* [55,56].

Results

SRS and IQ are well associated with ASD

Although the sample is enriched for ASD cases ($N = 35$ cases, 9.3% of the sample) relative to non-preterm cohorts, there is still a relatively low case-control ratio for a genome-wide study of this sample size (descriptive statistics for relevant covariates in **Table 1**). Therefore, we considered continuous measures of social impairment (SRS) and cognitive development (IQ) at age 10 for both associative and predictive analyses. Using continuous variables for SRS and IQ allow us to to study complexities beyond the ASD diagnostic categories [15,17]. **Figure 2A-B** shows the relationship between SRS, IQ, and ASD. The mean SRS is significantly higher in ASD cases compared to controls (mean difference of 1.74, $95\%\ CI$: $(1.41, 2.07)$). Mean IQ is significantly lower in ASD cases versus controls (mean difference of -2.23, $95\%\ CI$ $(-2.46, -1.96)$). Furthermore, SRS and IQ are negatively correlated (Pearson $\rho = -0.47, 95\%\ CI$: $(-0.55, -0.39)$). We also measured associations between demographic characteristics with SRS and IQ (**Figure 2C**) using multivariable regression. Male sex is associated with lower IQ, while public health insurance is associated with both lower IQ and increased social impairment. Demographic variables included in the multivariable regression explain approximately 12% and 15% of the total variance explained in IQ and SRS, as measured by adjusted R^2, with a summary of regression parameters in **Table 2**. Based on the associations identified here and the value of inclusion of continuous measures, subsequent genomic and epigenomic analyses control for demographic covariates.

Table 1: Descriptive statistics for demographi and clinical covariates

Continuous Variable	**Mean, SD, Median**
Maternal age	29.6, 6.61, 29.5
Gestational days	182.5, 9.17, 184.0
Categorical Variable	**Number (Proportion)**
ASD	
Case	35 (9.3%)
Control	344 (90.7%)
Race	
White	233 (61.5%)
Black	112 (29.5%)
Other	34 (9.0%)
Sex of baby	
Female	180 (47.5%)
Male	199 (52.5%)
Mother's smoking status	
Non-smoker	340 (89.7%)
Smoker	39 (10.3%)
Mother's insurance status	
Private	251 (66.2%)
Medicaid	128 (33.8%)

Table 2: Summary of regressions of SRS and IQ against clinical covariates.

	SRS		***IQ***	
Parameter	**Estimate (SE)**	**FDR-adjusted *P*-value (Raw *P*-value)**	**Estimate (SE)**	**FDR-adjusted *P*-value (Raw *P*-value)**
Race				
Black	0.219 (0.13)	0.165 (0.091)	-0.369 (0.13)	0.012 (0.004)
Other	0.375 (0.19)	0.087 (0.043)	-0.113 (0.18)	0.684 (0.533)
Sex				
Male	0.119 (0.10)	0.342 (0.243)	-0.288 (0.10)	0.012 (0.004)
Maternal age	-0.002 (0.01)	0.800 (0.800)	-0.003 (0.01)	0.792 (0.748)
Smoking status				
Yes	0.215 (0.17)	0.334 (0.204)	0.337 (0.17)	0.087 (0.043)
Mother's insurance				
Medicaid	0.454 (0.13)	0.002 (0.001)	-0.453 (0.13)	0.003 (0.001)
Gestational days	-0.017 (0.01)	0.012 (0.002)	0.012 (0.01)	0.087 (0.043)
Birthweight Z-score	-0.060 (0.05)	0.342 (0.247)	0.179 (0.05)	0.003 (0.001)
Placental inflammation	-0.042 (0.11)	0.793 (0.705)	-0.046 (0.11)	0.793 (0.677)

Genome-wide associations of mRNA, miRNA, and CpGs with SRS and IQ

Genome-wide association tests between each of the individual placental molecular datasets (e.g. the placental mRNA data, the CpG methylation, or the miRNA datasets) in relation to SRS and IQ (see

Methods) identified two genes with mRNA expression significantly associated with SRS at FDR-adjusted $P < 0.01$ (Hdc Homolog, Cell Cycle Regulator [*HECA*], LIM Domain Only 4 [*LMO4*]). We did not find CpG sites or miRNAs associated with SRS (**Table 3**). Associations between IQ and the mRNA expression, at FDR-adjusted $P < 0.01$, were observed at four genes, namely Ras-Related Protein Rab-5A (*RAB5A*), Transmembrane Protein 167A (*TMEM167A*), Signal Transducer and Activator of Transcription 2 (*STAT2*), ITPRIP Like 2 (*ITPRIPL2*). One CpG site (*cg09418354* located in the gene Carbohydrate Sulfotransferase 11 (*CHST11*) displayed an association with IQ, and no miRNAs were associated with IQ (**Table 3**). Manhattan plots (**Supplemental Figure 1**) show the strength of associations of all biomarkers by genomic position. Summary statistics for these associations are provided in **Supplemental Materials**. No mRNAs, CpG sites, or miRNAs were significantly associated with both SRS and IQ, though effect sizes for associations with the same features were in opposite directions (see **Supplemental Materials**).

Table 3: Summary of genome-wide associations of molecular profiles with SRS and IQ at FDR-adjusted $P < 0.01$.

SRS		
Biomarker	**Effect size**	**FDR-adjusted *P*-value**
mRNA expression		
HECA	0.571	0.001
LMO4	0.467	0.001
IQ		
Biomarker		
mRNA expression		
RAB5A	-0.516	0.002
TMEM167A	-0.632	0.004
ITPRIPL2	-0.557	0.004
STAT2	-0.584	0.004
CpG methylation site		
cg09418354	-0.005	0.002

Kernel regression shows predictive utility in aggregating multiple molecular datasets

Because the genome wide association analyses revealed few mRNAs, CpG sites or miRNAs that were associated with SRS or IQ with large effect sizes, we next assessed the impact of aggregating these molecular datasets on prediction of SRS and IQ. This was done to account for the considerable number

of biomarkers that have moderate effect sizes on outcome. To find the most parsimonious model with the greatest predictive performance, we first selected the optimal number of biomarkers per molecular profile for each outcome that gave the largest mean adjusted R^2 in predictive models with only one of the three molecular datasets (see **Supplemental Methods**). **Figure 3A** shows the relationship between the number of biomarkers from the mRNA expression, CpG level, miRNA expression datasets and their predictive performance. In general, predictive performance steadily increased as the number of biomarker features increased until reaching a tipping point where predictive performance decreased (**Figure 3A**). Overall, for CpG methylation, the top (lowest P-values of association) 5,000 CpG features showed the greatest predictive performance, and for the mRNA and miRNA expression datasets, the top 1,000 features showed the greatest predictive performance.

Using the fully-tuned 7,000 biomarkers (5,000 for CpG methylation and 1,000 for both mRNA and miRNA expression) per molecular dataset with feature selection done in the training set, we trained predictive models (both linear and Gaussian kernel models) using all individual, pair-wise, and triplet-wise combinations of the three molecular datasets. **Figure 3B** shows that whereas the mRNA had the lowest predicted performance to both IQ ($R^2 = 0.025$) and SRS ($R^2 = 0.025$), aggregating the mRNA expression, CpG methylation and miRNA expression datasets tends to increase the predictive performance. Specifically, in relation to both outcomes (SRS and IQ), the model using all three integrated datasets shows the greatest predictive performance (mean adjusted $R^2 = 0.11$ in IQ and $R^2 = 0.08$ in SRS).

Correlative networks of placental biomarkers

To gain further understanding of the associations among the identified mRNA, CpG and miRNA biomarkers in the context of IQ and SRS, we extracted ($n = 50$) mRNA, CpGs, and miRNAs that have the largest effect sizes on IQ and SRS in the kernel regression models and inferred sparse correlative networks using the graphical lasso [55,56] (see **Methods**). In the networks (**Supplemental Figure 2**), each molecular dataset clusters by itself, with minimal nodes extending between molecular datasets, and more interconnection is observed between miRNAs and CpG methylation versus mRNAs. These networks point to genes that play important roles in placental angiogenesis and neural function, such as

SMARCA2 (SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 2), *SLIT3* (Slit Guidance Ligand 3), and *LZTS2* (Leucine Zipper Tumor Suppressor 2) that have been previously associated with neurodevelopmental disorders, including intellectual disability, social impairment, mood disorders, and ASD [57–62].

Validation of in-sample and out-sample SRS and IQ prediction with ASD case and control

To contextualize our predictions, we tested whether the predicted SRS and IQ scores generated by our kernel models are associated with ASD case-control status; these predicted SRS and IQ scores represent the portion of the observed SRS and IQ values that our models can predict from placental genomic features. We used the optimal 7,000 biomarker features identified with a 10-fold cross-validation process, splitting samples into 10 hold-out sets and using the remaining samples as a training set to predict SRS and IQ for all 379 samples. After accounting for covariates, the predicted SRS and IQ values from the biomarker data were well-correlated with the observed clinical SRS and IQ values, explaining approximately 8% (approximate Spearman $\rho = 0.29$, cross-validatation R^2 P-value $P = 7.5 \times 10^{-9}$) and 12% (Spearman $\rho = 0.35, P = 3.6 \times 10^{-12}$) of the variance in the observed SRS and IQ variables, respectively. In addition, we found strong association between the predicted SRS and IQ with ASD case and controls, mean difference of -0.56 (test statistc $W = 8121, P = 6.6 \times 10^{-4}$) for IQ, and mean difference of 0.33 ($W = 4717, P = 0.03$) for SRS (**Figure 4**).

Because we lacked an external dataset with all three molecular data (mRNA, CpG methylation, and miRNA) and cognitive, social impairment and ASD data, we assessed the out-of-sample predictive performance of the CpG methylation-only models using MethylC-seq data from the MARBLES cohort (GEO GSE67615) [10]. We computed predicted IQ and SRS values for 47 placental samples (24 cases of ASD) and assessed differences in mean predicted IQ and SRS across ASD case and control groups. The direction of the association is similar to our data for IQ yet the differences in mean predicted IQ ($-0.22, P = 0.37$) and SRS ($-0.42, P = 0.12$) across ASD groups in MARBLES is not significant (**Figure 4**). This external validation provides some evidence of the portability of our models and merits further future validation of these models, as more placental multi-omic datasets are collected.

Discussion

We evaluated the predictive capability of three types of genomic and epigenomic molecular biomarkers (mRNA, CpG methylation, and miRNA) in the placenta on cognitive and social impairment in relation to ASD at 10 years of age. Genes that play important roles in placenta angiogenesis and neural function were associated with SRS and IQ. The multi-omic predictions of SRS and IQ are strong and explain up to 8% and 12% of the variance in the observed SRS and IQ variables in 5-fold cross-validation, respectively. This study supports the utility of aggregating information from biomarkers within and between molecular datasets to improve prediction of complex neurodevelopmental outcomes like social and intellectual ability, suggesting that traits on the placenta-brain axis may be omnigenic.

Several genes with known ties to neurodevelopmental disorders distinguished individuals with and without intellectual and social impairmenats. For example, CpG methylation in *SLIT3* was associated with intellectual (IQ) disability. *SLIT3* is highly expressed in trophoblastic endothelial cells [63] and plays a critical role in placental angiogenesis and in the development of neuronal connectivity. Human and animal genetic studies support that *SLIT3* is associated with mood disorders, IQ, and ASD [61,64–66]. *LZTS2*, another gene we found to be associated with IQ, is involved in regulating embryonic development by the *Wnt* signaling pathway [67,68]. Genetic and miRNA expression studies have linked *LZTS2* to social impairment and ASD [69–71]. Furthermore, *LZTS2* is bound by the Chromodomain Helicase DNA Binding Protein 8 gene (*CHD8*), which is associated with brain development in mice and neurodevelopmental disorders in humans [72–74]. In relation to social impairment, *ADAMTS6* was found to be associated with SRS.The *ADAMTS6* gene is a member of the ADAMTS protein family and is regulated by the cytokine TNF-alpha [75]. In previous studies, *ADAMTS6* has been implicated in intellectual disability and growth development and with socially affected traits in pigs [76,77].

Looking into the individual molecular datasets, DNA methylation effects showed the strongest prediction of both SRS and IQ impairment. There is strong evidence suggesting inverse correlation between DNA methylation of the first intron and gene expression across tissues and species [78]. We found that many

of the CpG loci with the largest effect sizes on SRS and IQ identified in our analysis are located near DNAase hyperactivity or active regulatory elements for the placenta [79,80], suggesting that these loci likely play regulatory functions. Experimental studies have demonstrated regions of the genome in which DNA methylation is causally important for gene regulation and those in which it is effectively silent [81]. We found that aggregating biomarkers within and between molecular datasets improves prediction of social and cognitive impairment. Specifically, this observation suggests new possibilities to the discovery of candidate genes in the placenta that convey neurodevelopmental risk, improving the understanding of the placenta-brain axis. Recent work in transcriptome-wide association studies (TWAS) are a promising tool that aggregates genetics and transcriptomics to identify candidate trait-associated genes [82,83]. Incorporating information from regulatory biomarkers, like transcription factors and miRNAs, into TWAS increases study power to generate hypotheses about regulation [84,85]. Given our observations in this analysis and the number of the integrated molecular datasets, we believe that the ELGAN study can be used to train predictive models for placental transcriptomics from genetics, enriched for regulatory elements [85]. These transcriptomic models can then be applied to genome-wide association study cohorts to study the regulation of gene-trait associations in the placenta.

Limitations

When interpreting the results of this study, some factors should be considered. Extremely preterm birth is strongly associated with increased risk for neurodevelopmental disorders [18]. This association may lead to bias in estimated associations between the molecular biomarkers and outcomes, especially when unmeasured confounders are linked to both pre-term birth and autism [86]. Still, to our knowledge the ELGAN cohort is currently the largest available placental repository with both multiple molecular datasets and long-term neurodevelopmental assessment of the children. Second, as the placenta is comprised of several heterogeneous cell types, tissue-specific molecular patterns in the placenta should be taken into consideration when interpreting these findings in relation to other tissue samples; future comparison between tissues will not be straightforward. Lastly, to test the reproducibility and robustness of our kernel models, we believe further out-of-sample validation is required, using datasets with larger sample sizes and similar molecular datasets. Though in-sample predictive performance is strong, platform differences

between the ELGAN training set (assayed with EPIC BeadChip) and validation set (MethylC-seq) may lead to loss of predictive power. As our optimal models all aggregate various datasets, the dearth of data for the placenta, in the context of social and intellectual impairment, makes out-of-sample validation especially challenging. Lack of external validation may render our analysis exploratory in nature, but we provide evidence of a link between molecular features within the fetal placenta and social and cognitive outcomes in children that merits future investigation.

Conclusions

Our analysis underscores the importance of synthesizing data representing various levels of biological data to understand distinct genomic and epigenomc underpinnings of complex developmental deficits, like intellectual and social impairment. This study provides novel evidence for the omnigenicity of the placenta-brain axis in the context of social and intellectual impairment.

List of Abbreviations

Extremely Low Gestational Age Newborn (ELGAN)

Intellectual ability (IQ)

Differential Ability Scales-II (DAS-II)

Social Responsiveness Scale (SRS)

Autism Spectrum Disorder (ASD)

SRS gender-normed T-score (SRS-T)

Epigenome-wide association study (EWAS)

Markers of Autism Risk in Babies-Learning Early Signs (MARBLES)

Kernel regression least squares (KRLS)

Declarations

Ethics approval and consent to participate

The study was approved by the Institutional Review Board of the University of North Carolina at Chapel Hill. All participants consented to the study as per IRB protocol.

Consent for publication

Not applicable

Availability of data and materials

Multiomic data from the ELGAN study is available from the NCBI Gene Expression Omnibus GSE154829. All genomic and clinical data is also available upon request to H.P.S. For validation, we used MethylC-seq data from the MARBLES study available at GSE67615.

Competing interests

The authors have no competing financial interests to disclose.

Funding

This study was supported by grants from the National Institutes of Health (NIH), specifically the National Institute of Neurological Disorders and Stroke (U01NS040069; R01NS040069), the Office of the NIH Director (UG3OD023348), the National Institute of Environmental Health Sciences (T32-ES007018), National Institute of Nursing Research (K23NR017898), and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01HD092374).

Authors' Contributions

H.S.P, R.M.J., L.S., K.C.K.K., C.J.M., T.M.O, and R.C.F. conceived and designed the study. H.S.P. and A.B. acquired and analyzed the data. H.S.P., A.B. and R.C.F. interpreted data. H.S.P. and A.B. drafted the work and all authors revised. All authors have approved the submitted version and are accountable for their own contributions.

Acknowledgements

We would like to thank the study participants of the ELGAN-ECHO study. We would also like to thank Michael Love and Lana Garmire for helpful discussion during the research process.

References

1. Hodyl NA, Aboustate N, Bianco-Miotto T, Roberts CT, Clifton VL, Stark MJ. Child neurodevelopmental outcomes following preterm and term birth: What can the placenta tell us? Placenta. W.B. Saunders Ltd; 2017. p. 79–86.

2. Hu WF, Chahrour MH, Walsh CA. The Diverse Genetic Landscape of Neurodevelopmental Disorders. Annu Rev Genomics Hum Genet. Annual Reviews; 2014;15:195–213.

3. Agrawal S, Rao SC, Bulsara MK, Patole SK. Prevalence of autism spectrum disorder in preterm infants: A meta-Analysis. Pediatrics. American Academy of Pediatrics; 2018.

4. Xie S, Heuvelman H, Magnusson C, Rai D, Lyall K, Newschaffer CJ, et al. Prevalence of Autism Spectrum Disorders with and without Intellectual Disability by Gestational Age at Birth in the Stockholm Youth Cohort: a Register Linkage Study. Paediatr Perinat Epidemiol. Blackwell Publishing Ltd; 2017;31:586–94.

5. Korzeniewski SJ, Joseph RM, Kim SH, Allred EN, O'Shea TM, Leviton A, et al. Social Responsiveness Scale Assessment of the Preterm Behavioral Phenotype in 10-Year-Olds Born Extremely Preterm. J Dev Behav Pediatr [Internet]. 2017 [cited 2019 Sep 5];38:697–705. Available from: http://insights.ovid.com/crossref?an=00004703-201711000-00003

6. Rosenfeld CS. The placenta-brain-axis. J. Neurosci. Res. John Wiley and Sons Inc.; 2020.

7. Rosenfeld CS. Placental serotonin signaling, pregnancy outcomes, and regulation of fetal brain development†. Biol Reprod. Oxford Academic; 2020;102:532–8.

8. Meakin CJ, Martin EM, Santos HP, Mokrova I, Kuban K, O'Shea TM, et al. Placental CpG methylation of HPA-axis genes is associated with cognitive impairment at age 10 among children born extremely preterm. Horm Behav. Academic Press Inc.; 2018;101:29–35.

9. Paquette AG, Houseman EA, Green BB, Lesseur C, Armstrong DA, Lester B, et al. Regions of variable DNA methylation in human placenta associated with newborn neurobehavior. Epigenetics [Internet]. 2016 [cited 2019 Sep 30];11:603–13. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27366929

10. Schroeder DI, Schmidt RJ, Crary-Dooley FK, Walker CK, Ozonoff S, Tancredi DJ, et al. Placental methylome analysis from a prospective autism study. Mol Autism [Internet]. 2016 [cited 2019 Sep 29];7:51. Available from: http://molecularautism.biomedcentral.com/articles/10.1186/s13229-016-0114-8

480 11. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic
481 risk variants for autism spectrum disorder. Nat Genet [Internet]. NIH Public Access; 2019 [cited 2019 Aug
482 30];51:431–44. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30804558
483 12. Sullivan PF, Agrawal A, Bulik CM, Andreassen OA, Børglum AD, Breen G, et al. Psychiatric
484 Genomics: An Update and an Agenda. Am J Psychiatry [Internet]. NIH Public Access; 2018 [cited 2019
485 Aug 30];175:15. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28969442
486 13. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic.
487 Cell [Internet]. Elsevier; 2017 [cited 2019 Jun 7];169:1177–86. Available from:
488 http://www.ncbi.nlm.nih.gov/pubmed/28622505
489 14. Liu X, Li YI, Pritchard JK. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. Cell.
490 Cell Press; 2019;177:1022-1034.e6.
491 15. Geschwind DH, State MW. Gene hunting in autism spectrum disorder: On the path to precision
492 medicine [Internet]. Lancet Neurol. Lancet Publishing Group; 2015 [cited 2020 Aug 20]. p. 1109–20.
493 Available from: /pmc/articles/PMC4694565/?report=abstract
494 16. O'Shea TM, Allred EN, Dammann O, Hirtz D, Kuban KCK, Paneth N, et al. The ELGAN study of the
495 brain and related disorders in extremely low gestational age newborns. Early Hum Dev. 2009;85:719–25.
496 17. Torske T, Nærland T, Bettella F, Bjella T, Malt E, Høyland AL, et al. Autism spectrum disorder
497 polygenic scores are associated with every day executive function in children admitted for clinical
498 assessment. Autism Res [Internet]. John Wiley and Sons Inc.; 2020 [cited 2020 Aug 20];13:207–20.
499 Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/aur.2207
500 18. Joseph RM, O'Shea TM, Allred EN, Heeren T, Hirtz D, Paneth N, et al. Prevalence and associated
501 features of autism spectrum disorder in extremely low gestational age newborns at age 10 years. Autism
502 Res [Internet]. John Wiley & Sons, Ltd; 2017 [cited 2019 Sep 5];10:224–32. Available from:
503 http://www.ncbi.nlm.nih.gov/pubmed/27220677
504 19. Joseph RM, O'Shea TM, Allred EN, Heeren T, Hirtz D, Jara H, et al. Neurocognitive and Academic
505 Outcomes at Age 10 Years of Extremely Preterm Newborns. Pediatrics [Internet]. American Academy of
506 Pediatrics; 2016 [cited 2019 Sep 5];137:e20154343. Available from:
507 http://www.ncbi.nlm.nih.gov/pubmed/27006473

508 20. Beran TN. Elliott, C. D. (2007). Differential Ability Scales (2nd ed.). San Antonio, TX: Harcourt
509 Assessment. Can J Sch Psychol [Internet]. Sage PublicationsSage CA: Los Angeles, CA; 2007 [cited
510 2019 Sep 5];22:128–32. Available from: http://journals.sagepub.com/doi/10.1177/0829573507302967
511 21. Constantino JN, Davis SA, Todd RD, Schindler MK, Gross MM, Brophy SL, et al. Validation of a brief
512 quantitative measure of autistic traits: comparison of the social responsiveness scale with the autism
513 diagnostic interview-revised. J Autism Dev Disord [Internet]. 2003 [cited 2019 Sep 5];33:427–33.
514 Available from: http://www.ncbi.nlm.nih.gov/pubmed/12959421
515 22. Constantino JN, Zhang Y, Frazier T, Abbacchi AM, Law P. Sibling Recurrence and the Genetic
516 Epidemiology of Autism. Am J Psychiatry [Internet]. 2010 [cited 2019 Sep 5];167:1349–56. Available
517 from: http://www.ncbi.nlm.nih.gov/pubmed/20889652
518 23. Rutter, Bailey, Lord. The social communication questionnare. Western Ps. Los Angeles; 2003.
519 24. Risi S, Lord C, Gotham K, Corsello C, Chrysler C, Szatmari P, et al. Combining information from
520 multiple sources in the diagnosis of autism spectrum disorders. J Am Acad Child Adolesc Psychiatry
521 [Internet]. Elsevier; 2006 [cited 2019 Sep 5];45:1094–103. Available from:
522 http://www.ncbi.nlm.nih.gov/pubmed/16926617
523 25. Lord C, Rutter M, DiLavore P, Risi S, Gotham K, Bishop S. Autism Diagnostic Observation Schedule.
524 Los Angeles: Western Psychological Corporation; 2012.
525 26. Addo KA, Bulka C, Dhingra R, Santos HP, Jr, Smeester L, et al. Acetaminophen use during
526 pregnancy and DNA methylation in the placenta of the extremely low gestational age newborn (ELGAN)
527 cohort. Environ Epigenetics. Oxford University Press; 2019;5.
528 27. Santos HP, Bhattacharya A, Martin EM, Addo K, Psioda M, Smeester L, et al. Epigenome-wide DNA
529 methylation in placentas from preterm infants: association with maternal socioeconomic status.
530 Epigenetics [Internet]. Taylor and Francis Inc.; 2019 [cited 2019 Aug 4];14:751–65. Available from:
531 http://www.ncbi.nlm.nih.gov/pubmed/31062658
532 28. Bulka CM, Dammann O, Santos HP, VanderVeen DK, Smeester L, Fichorova R, et al. Placental CpG
533 methylation of inflammation, angiogenic, and neurotrophic genes and retinopathy of prematurity. Investig
534 Ophthalmol Vis Sci. Association for Research in Vision and Ophthalmology Inc.; 2019;60:2888–94.
535 29. Clark J, Martin E, Bulka CM, Smeester L, Santos HP, O'Shea TM, et al. Associations between

placental CpG methylation of metastable epialleles and childhood body mass index across ages one, two and ten in the Extremely Low Gestational Age Newborns (ELGAN) cohort. Epigenetics. Taylor and Francis Inc.; 2019;14:1102–11.

30. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics [Internet]. 2014 [cited 2019 Aug 4];30:1363–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24478339

31. Fortin J-P, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol [Internet]. 2014 [cited 2019 Aug 4];15:503. Available from: http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0503-2

32. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics [Internet]. 2007 [cited 2019 Aug 4];8:118–27. Available from: https://academic.oup.com/biostatistics/article/8/1/118/252073

33. Triche TJ, Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. Nucleic Acids Res [Internet]. Oxford University Press; 2013 [cited 2020 Aug 14];41:e90. Available from: /pmc/articles/PMC3627582/?report=abstract

34. Fortin JP, Triche TJ, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Bioinformatics [Internet]. Oxford University Press; 2017 [cited 2020 Aug 14];33:558–60. Available from: /pmc/articles/PMC5408810/?report=abstract

35. Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics [Internet]. 2010 [cited 2019 Dec 4];11:587. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-11-587

36. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods [Internet]. 2017 [cited 2019 Aug 5];14:417–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28263959

37. Qi Z, Wang L, Desai K, Cogswell J, Stern M, Lawson B, et al. Reliable Gene Expression Profiling from

564 Small and Hematoxylin and Eosin–Stained Clinical Formalin-Fixed, Paraffin-Embedded Specimens Using
565 the HTG EdgeSeq Platform. J Mol Diagnostics. Elsevier B.V.; 2019;21:796–807.
566 38. Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and
567 differential expression in mRNA-Seq experiments. BMC Bioinformatics [Internet]. BioMed Central; 2010
568 [cited 2019 Mar 22];11:94. Available from:
569 https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-11-94
570 39. Risso D, Ngai J, Speed TP, Dudoit S. Normalization of RNA-seq data using factor analysis of control
571 genes or samples. Nat Biotechnol [Internet]. Nature Publishing Group; 2014 [cited 2019 Aug 5];32:896–
572 902. Available from: http://www.nature.com/articles/nbt.2931
573 40. Gagnon-Bartsch JA, Speed TP. Using control genes to correct for unwanted variation in microarray
574 data. Biostatistics [Internet]. Oxford University Press; 2012 [cited 2019 Sep 15];13:539–52. Available
575 from: http://www.ncbi.nlm.nih.gov/pubmed/22101192
576 41. Risso D, Ngai J, Speed TP, Dudoit S. Normalization of RNA-seq data using factor analysis of control
577 genes or samples. Nat Biotechnol [Internet]. Nature Publishing Group; 2014 [cited 2019 Mar 22];32:896–
578 902. Available from: http://www.nature.com/articles/nbt.2931
579 42. Phipson B, Lee S, Majewski IJ, Alexander WS, Smyth GK. ROBUST HYPERPARAMETER
580 ESTIMATION PROTECTS AGAINST HYPERVARIABLE GENES AND IMPROVES POWER TO
581 DETECT DIFFERENTIAL EXPRESSION. Ann Appl Stat [Internet]. NIH Public Access; 2016 [cited 2019
582 Aug 5];10:946–63. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28367255
583 43. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data
584 with DESeq2. Genome Biol [Internet]. BioMed Central; 2014 [cited 2019 Mar 22];15:550. Available from:
585 http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8
586 44. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to
587 Multiple [Internet]. Source J. R. Stat. Soc. Ser. B. 1995. Available from:
588 https://www.jstor.org/stable/pdf/2346101.pdf?refreqid=excelsior%3A6411207ed4b0feb82c3964cc8b8151
589 cb
590 45. Tsai P-C, Bell JT. Power and sample size estimation for epigenome-wide association scans to detect
591 differential DNA methylation. Int J Epidemiol [Internet]. 2015 [cited 2019 Aug 5];44:1429–41. Available

592 from: https://academic.oup.com/ije/article-lookup/doi/10.1093/ije/dyv041

593 46. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression

594 analyses for RNA-sequencing and microarray studies. Nucleic Acids Res [Internet]. Oxford University

595 Press; 2015 [cited 2019 Aug 5];43:e47–e47. Available from:

596 http://www.ncbi.nlm.nih.gov/pubmed/25605792

597 47. Morota G, Gianola D. Kernel-based whole-genome prediction of complex traits: a review. Front Genet

598 [Internet]. Frontiers Media SA; 2014 [cited 2019 Aug 30];5:363. Available from:

599 http://www.ncbi.nlm.nih.gov/pubmed/25360145

600 48. Zhu B, Song N, Shen R, Arora A, Machiela MJ, Song L, et al. Integrating Clinical and Multiple Omics

601 Data for Prognostic Assessment across Human Cancers. Sci Rep [Internet]. Nature Publishing Group;

602 2017 [cited 2019 Aug 27];7:16954. Available from: http://www.nature.com/articles/s41598-017-17031-8

603 49. Elisseeff A, Technologies B, Shawe-Taylor J, Holloway R, Kandola J. On Kernel-Target Alignment N

604 ello Cristianini BIOwulf Technologies nello@support-vector. net [Internet]. Available from:

605 https://papers.nips.cc/paper/1946-on-kernel-target-alignment.pdf

606 50. Endelman JB. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP.

607 Plant Genome. 2011;4:250–5.

608 51. Zhu W, Hu B, Becker C, Doğan ES, Berendzen KW, Weigel D, et al. Altered chromatin compaction

609 and histone methylation drive non-additive gene expression in an interspecific Arabidopsis hybrid.

610 Genome Biol [Internet]. BioMed Central Ltd.; 2017 [cited 2020 Aug 14];18:1–16. Available from:

611 https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1281-4

612 52. Zeng Y, Amador C, Xia C, Marioni R, Sproul D, Walker RM, et al. Parent of origin genetic effects on

613 methylation in humans are common and influence complex trait variation. Nat Commun [Internet]. Nature

614 Publishing Group; 2019 [cited 2020 Aug 14];10:1–13. Available from: https://doi.org/10.1038/s41467-019-

615 09301-y

616 53. Hainmueller J, Hazlett C. Kernel Regularized Least Squares: Reducing Misspecification Bias with a

617 Flexible and Interpretable Machine Learning Approach. Polit Anal [Internet]. 2014 [cited 2019 Sep

618 5];22:143–68. Available from: http://pan.oxfordjournals.org/

619 54. Xu QS, Liang YZ. Monte Carlo cross validation. Chemom Intell Lab Syst. Elsevier; 2001;56:1–11.

620 55. Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso.
621 Biostatistics. 2007;1–10.
622 56. Epskamp S, Cramer AOJ, Waldorp LJ, Schmittmann VD, Borsboom D. Qgraph: Network
623 visualizations of relationships in psychometric data. J Stat Softw. 2012;48.
624 57. Crawley JN, Heyer WD, LaSalle JM. Autism and Cancer Share Risk Genes, Pathways, and Drug
625 Targets. Trends Genet. Elsevier Ltd; 2016. p. 139–46.
626 58. Chater-Diehl E, Ejaz R, Cytrynbaum C, Siu MT, Turinsky A, Choufani S, et al. New insights into DNA
627 methylation signatures: SMARCA2 variants in Nicolaides-Baraitser syndrome. BMC Med Genomics
628 [Internet]. BioMed Central Ltd.; 2019 [cited 2020 Apr 22];12:105. Available from:
629 https://bmcmedgenomics.biomedcentral.com/articles/10.1186/s12920-019-0555-y
630 59. Tang S, Hughes E, Lascelles K, Simpson MA, Pal DK, Marini C, et al. New SMARCA2 mutation in a
631 patient with Nicolaides–Baraitser syndrome and myoclonic astatic epilepsy. Am J Med Genet Part A.
632 Wiley-Liss Inc.; 2017;173:195–9.
633 60. Koga M, Ishiguro H, Yazaki S, Horiuchi Y, Arai M, Niizato K, et al. Involvement of SMARCA2/BRM in
634 the SWI/SNF chromatin-remodeling complex in schizophrenia. Hum Mol Genet [Internet]. 2009 [cited
635 2020 Apr 25];18:2483–94. Available from: http://snp.ims.u-tokyo.ac.jp/cgi-bin/SnpInfo.cgi?SNP_ID=IMS-
636 JST050328.
637 61. Glessner JT, Wang K, Sleiman PMA, Zhang H, Kim CE, Flory JH, et al. Duplication of the slit3 locus
638 on 5q35.1 predisposes to major depressive disorder. PLoS One. Public Library of Science; 2010;5.
639 62. Thiffault I, Zuccarelli B, Welsh H, Yuan X, Farrow E, Zellmer L, et al. Hypotonia and intellectual
640 disability without dysmorphic features in a patient with PIGN-related disease. BMC Med Genet [Internet].
641 BioMed Central Ltd.; 2017 [cited 2020 Apr 25];18:124. Available from:
642 http://bmcmedgenet.biomedcentral.com/articles/10.1186/s12881-017-0481-9
643 63. Liao W-X, Laurent LC, Agent S, Hodges J, Chen D. Human Placental Expression of SLIT/ROBO
644 Signaling Cues: Effects of Preeclampsia and Hypoxia1. Biol Reprod. Oxford University Press (OUP);
645 2012;86.
646 64. Cukier HN, Dueker ND, Slifer SH, Lee JM, Whitehead PL, Lalanne E, et al. Exome sequencing of
647 extended families with autism reveals genes shared across neurodevelopmental and neuropsychiatric

disorders. Mol Autism [Internet]. BioMed Central; 2014 [cited 2020 Apr 22];5:1. Available from: http://molecularautism.biomedcentral.com/articles/10.1186/2040-2392-5-1

65. Christoforou A, Espeseth T, Davies G, Fernandes CPD, Giddaluru S, Mattheisen M, et al. GWAS-based pathway analysis differentiates between fluid and crystallized intelligence. Genes, Brain Behav. Blackwell Publishing Ltd; 2014;13:663–74.

66. Park SM, Plachez C, Huang S. Sex-dependent motor deficit and increased anxiety-like states in mice lacking autism-associated gene slit3. Front Behav Neurosci. Frontiers Media S.A.; 2018;12.

67. Stalman SE, Solanky N, Ishida M, Alemán-Charlet C, Abu-Amero S, Alders M, et al. Genetic Analyses in Small-for-Gestational-Age Newborns. J Clin Endocrinol Metab [Internet]. Oxford University Press; 2018 [cited 2020 Apr 22];103:917–25. Available from: https://academic.oup.com/jcem/article/103/3/917/4801232

68. Thyssen G, Li T-H, Lehmann L, Zhuo M, Sharma M, Sun Z. LZTS2 Is a Novel β-Catenin-Interacting Protein and Regulates the Nuclear Export of β-Catenin. Mol Cell Biol. American Society for Microbiology; 2006;26:8857–67.

69. Chang Y, Lin C, Huang H, Chang J, Kuo H. Chromosomal microarray and whole-exome sequence analysis in Taiwanese patients with autism spectrum disorder. Mol Genet Genomic Med [Internet]. Wiley-Blackwell; 2019 [cited 2020 Apr 22];7. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/mgg3.996

70. Connolly JJ, Glessner JT, Hakonarson H. A Genome-Wide Association Study of Autism Incorporating Autism Diagnostic Interview-Revised, Autism Diagnostic Observation Schedule, and Social Responsiveness Scale. Child Dev. Child Dev; 2013;84:17–33.

71. Lackinger M, Sungur AÖ, Daswani R, Soutschek M, Bicker S, Stemmler L, et al. A placental mammal-specific micro RNA cluster acts as a natural brake for sociability in mice. EMBO Rep [Internet]. EMBO; 2019 [cited 2020 Apr 30];20. Available from: https://onlinelibrary.wiley.com/doi/abs/10.15252/embr.201846429

72. Gompers AL, Su-Feher L, Ellegood J, Copping NA, Riyadh MA, Stradleigh TW, et al. Germline Chd8 haploinsufficiency alters brain development in mouse. Nat Neurosci. Nature Publishing Group; 2017;20:1062–73.

676 73. Sugathan A, Biagioli M, Golzio C, Erdin S, Blumenthal I, Manavalan P, et al. CHD8 regulates
677 neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. Proc Natl
678 Acad Sci U S A. National Academy of Sciences; 2014;111:E4468–77.
679 74. Yasin H, Gibson WT, Langlois S, Stowe RM, Tsang ES, Lee L, et al. A distinct neurodevelopmental
680 syndrome with intellectual disability, autism spectrum disorder, characteristic facies, and macrocephaly is
681 caused by defects in CHD8. J Hum Genet. Nature Publishing Group; 2019;64:271–80.
682 75. Kelwick R, Desanlis I, Wheeler GN, Edwards DR. The ADAMTS (A Disintegrin and Metalloproteinase
683 with Thrombospondin motifs) family. Genome Biol [Internet]. BioMed Central Ltd.; 2015 [cited 2020 May
684 12];16:113. Available from: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-015-0676-
685 3
686 76. Malli T, Duba HC, Erdel M, Marschon R, Kranewitter W, Deutschbauer S, et al. Disruption of the
687 ARID1B and ADAMTS6 loci due to a t(5;6)(q12.3;q25.3) in a patient with developmental delay. Am J Med
688 Genet Part A [Internet]. Wiley-Liss Inc.; 2014 [cited 2020 May 12];164:3126–31. Available from:
689 http://www.ncbi.nlm.nih.gov/pubmed/25250687
690 77. Wu P, Wang K, Yang Q, Zhou J, Chen D, Liu Y, et al. Whole-genome re-sequencing association
691 study for direct genetic effects and social genetic effects of six growth traits in Large White pigs. Sci Rep.
692 Nature Publishing Group; 2019;9:1–12.
693 78. Anastasiadi D, Esteve-Codina A, Piferrer F. Consistent inverse correlation between DNA methylation
694 of the first intron and gene expression across tissues and species. Epigenetics Chromatin [Internet]. 2018
695 [cited 2020 Jan 19];11:37. Available from:
696 https://epigeneticsandchromatin.biomedcentral.com/articles/10.1186/s13072-018-0205-1
697 79. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al.
698 Integrative analysis of 111 reference human epigenomes. Nature. Nature Publishing Group;
699 2015;518:317–29.
700 80. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. The Encyclopedia of DNA
701 elements (ENCODE): Data portal update. Nucleic Acids Res. Oxford University Press; 2018;46:D794--
702 D801.
703 81. Lea AJ, Vockley CM, Johnston RA, Del Carpio CA, Barreiro LB, Reddy TE, et al. Genome-wide

quantification of the effects of DNA methylation on human gene regulation. Elife. eLife Sciences Publications Ltd; 2018;7.

82. Gamazon ER, Wheeler HE, Shah KP, Mozaffari S V, Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet [Internet]. Nature Publishing Group; 2015 [cited 2019 Apr 6];47:1091–8. Available from: http://www.nature.com/articles/ng.3367

83. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet [Internet]. 2016 [cited 2019 Apr 6];48:245–52. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26854917

84. Zhang W, Voloudakis G, Rajagopal VM, Readhead B, Dudley JT, Schadt EE, et al. Integrative transcriptome imputation reveals tissue-specific and shared biological mechanisms mediating susceptibility to complex traits. Nat Commun [Internet]. Nature Publishing Group; 2019 [cited 2019 Oct 2];10:3834. Available from: http://www.nature.com/articles/s41467-019-11874-7

85. Bhattacharya A, Love MI. Multi-omic strategies for transcriptome-wide prediction and association studies. bioRxiv. Cold Spring Harbor Laboratory; 2020;2020.04.17.047225.

86. Paternoster L, Tilling K, Davey Smith G. Genetic epidemiology and Mendelian randomization for informing disease therapeutics: Conceptual and methodological challenges. Barsh GS, editor. PLOS Genet [Internet]. Public Library of Science; 2017 [cited 2019 Jul 15];13:e1006944. Available from: https://dx.plos.org/10.1371/journal.pgen.1006944

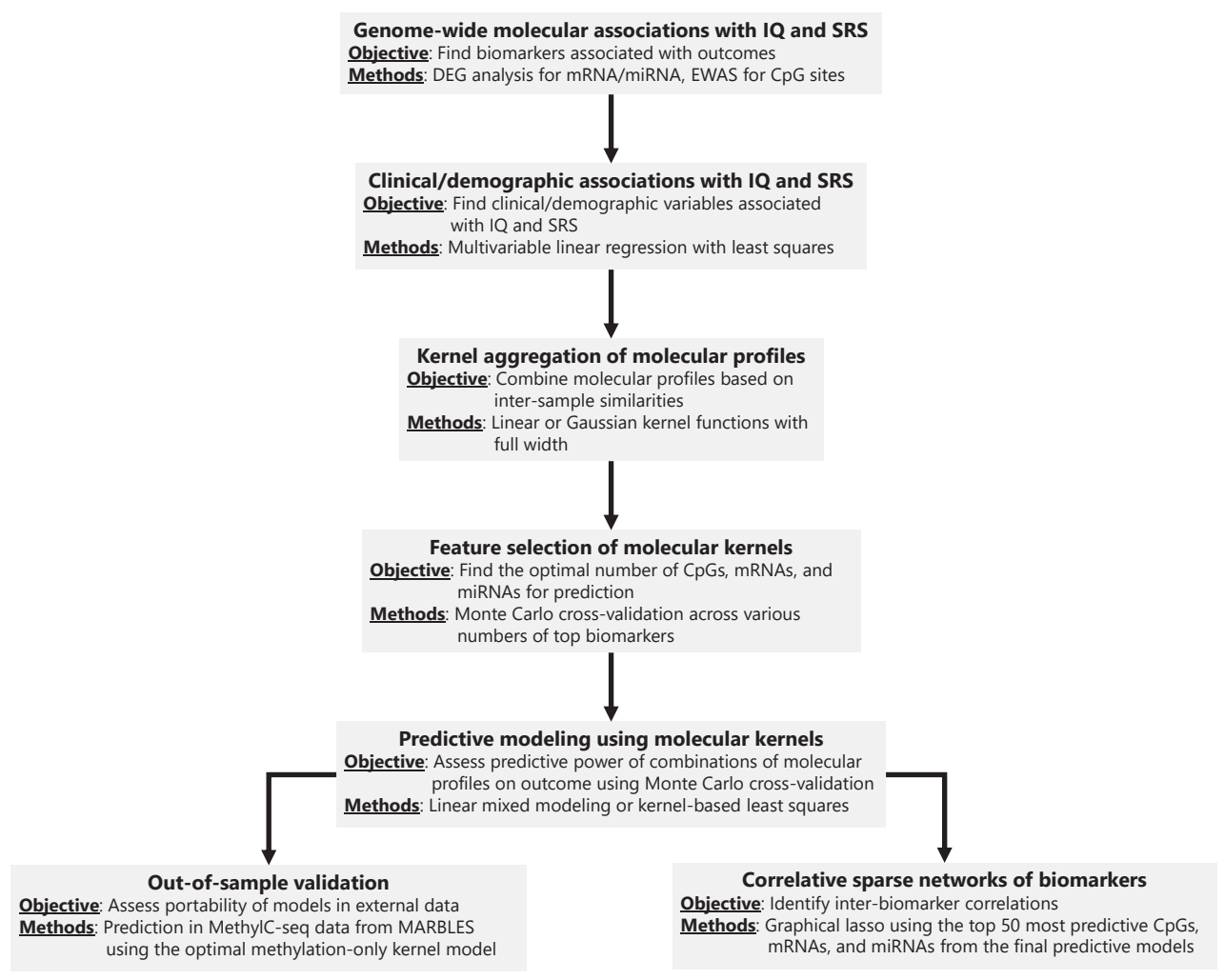
Figure Captions

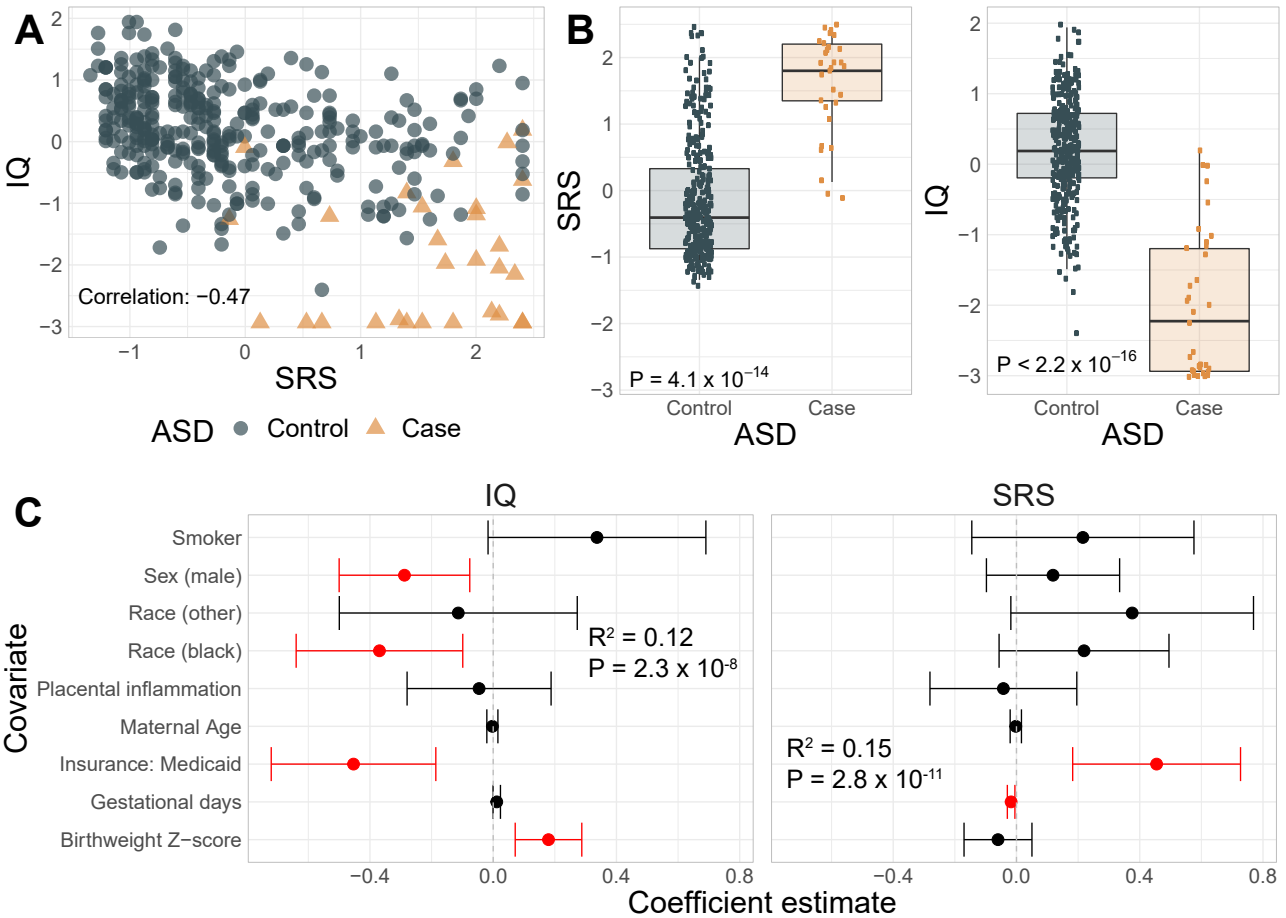
Figure 1: ***Scheme for kernel aggregation and prediction models****.* (1) Design matrices for CpG sites, mRNAs, and miRNAs are aggregated to form a linear or Gaussian kernel matrix that measures the similarity of samples. (2) Clinical variables are regressed out of the outcomes IQ and SRS and from the omic kernels to limit influence from these variables. (3) Using 50-fold Monte Carlo cross-validation on 75%-25% training-test splits, we train prediction models with the kernel matrices for IQ and SRS in the training set and predict in the test sets. Prediction is assessed in every fold with adjusted R^2 and averaged for an overall prediction metric.

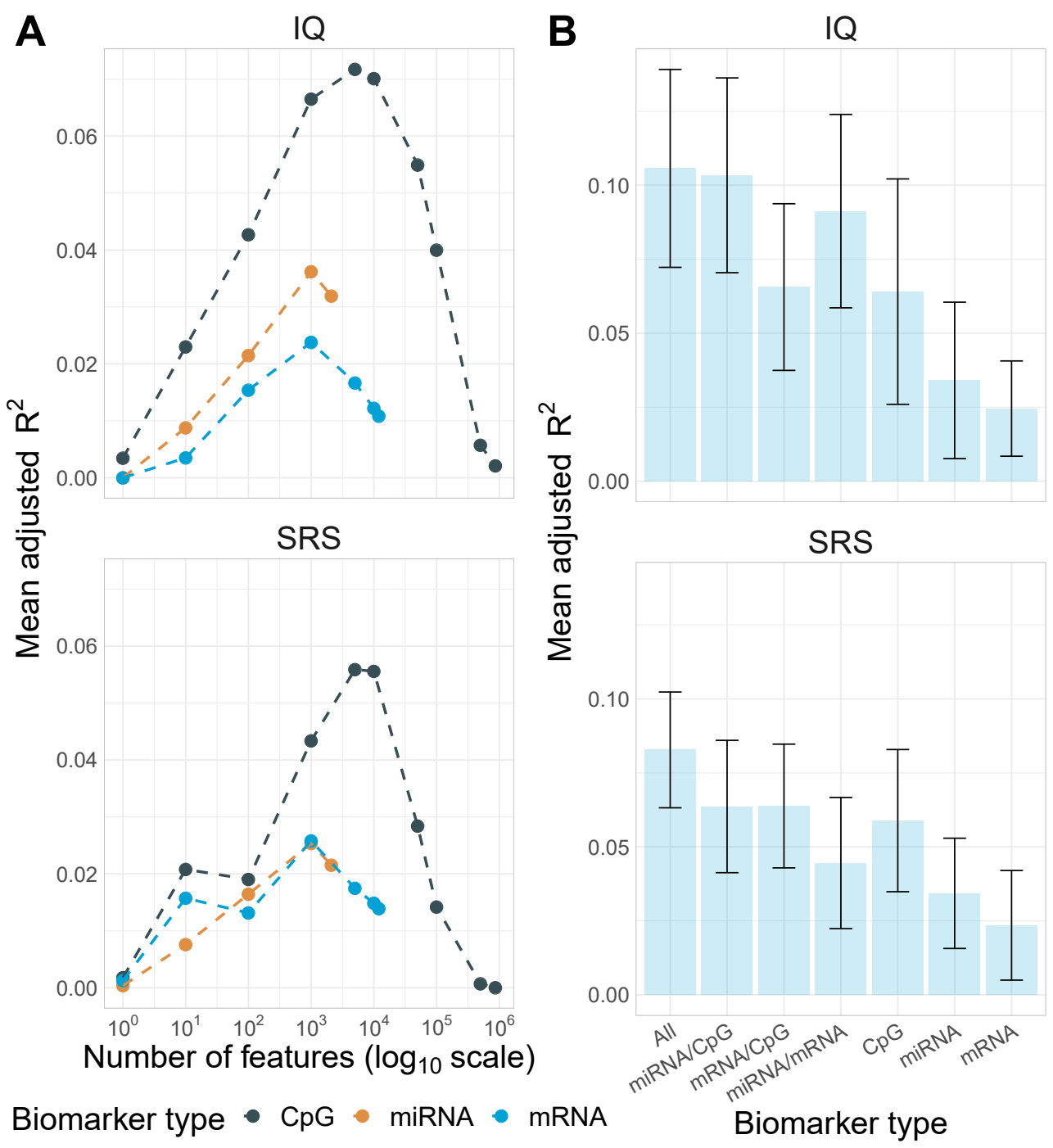
Figure 2: ***Associations between SRS, IQ, and ASD and with clinical variables.*** **(A)** Scatter plot of SRS (X-axis) and IQ (Y-axis) colored by ASD case (orange) and control (blue) status. **(B)** Boxplots of SRS and IQ across ASD case-control status. *P*-value from a two-sample Mann-Whitney test is provided. **(C)** Caterpillar plot of multivariable linear regression parameters of IQ and SRS using clinical variables. Points give the regression parameter estimates with error bars showing the 95% FDR-adjusted confidence intervals [44]. The null value of 0 is provided for reference with the dotted line.

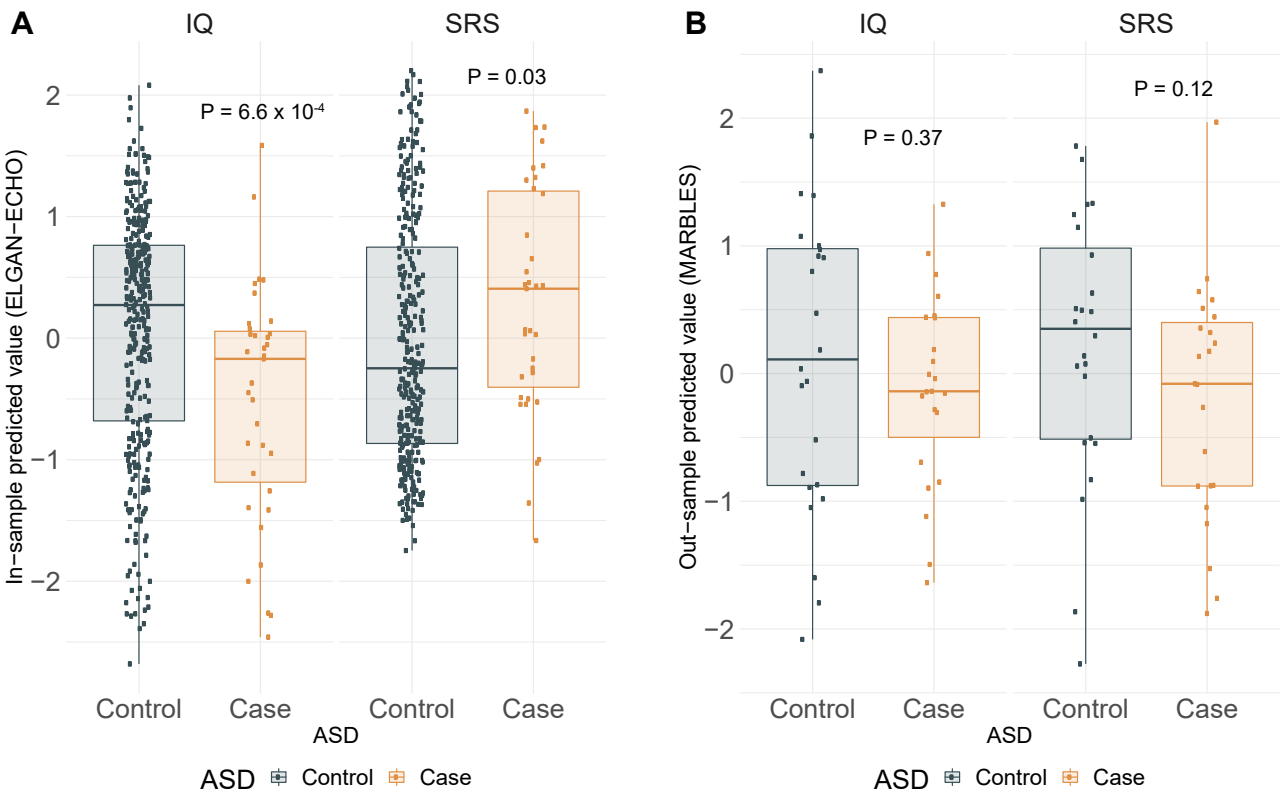
Figure 3: ***In-sample predictive performance of kernel models****.* (A) Adjusted mean R^2 (Y-axis) of best kernel models over various numbers of the top biomarkers (X-axis) in the CpG (dark blue), miRNA (orange), and mRNA (light blue) omics over 50 Monte Carlo folds. The X-axis scale is logarithmic. (B) Bar plots of adjusted mean R^2 (Y-axis) for optimally tuned kernel predictive models using all combinations of omics (X-axis) over 50 Monte Carlo folds. The error bar gives a spread of one standard deviation around the mean adjusted R^2.

Figure 4: ***Association of ASD case/control status with predicted SRS and IQ*****.** **(A)** Box-plots of in-sample predicted IQ (left) and SRS (right) over ASD case/control in ELGAN over 10-fold cross-validation. **(B)** Box-plots of out-sample predicted IQ (left) and SRS (right) over ASD case/control in MARBLES external validation dataset. P-values presented as from a Mann-Whitney test of differences across the ASD case/control groups.


Genome-wide molecular associations with IQ and SRS
Objective: Find biomarkers associated with outcomes
Methods: DEG analysis for mRNA/miRNA, EWAS for CpG sites
Clinical/demographic associations with IQ and SRS
Objective: Find clinical/demographic variables associated with IQ and SRS
Methods: Multivariable linear regression with least squares
Kernel aggregation of molecular profiles
Objective: Combine molecular profiles based on inter-sample similarities
Methods: Linear or Gaussian kernel functions with full width
Feature selection of molecular kernels
Objective: Find the optimal number of CpGs, mRNAs, and miRNAs for prediction
Methods: Monte Carlo cross-validation across various numbers of top biomarkers
Predictive modeling using molecular kernels
Objective: Assess predictive power of combinations of molecular profiles on outcome using Monte Carlo cross-validation
Methods: Linear mixed modeling or kernel-based least squares
Out-of-sample validation
Objective: Assess portability of models in external data
Methods: Prediction in MethylC-seq data from MARBLES using the optimal methylation-only kernel model
Correlative sparse networks of biomarkers
Objective: Identify inter-biomarker correlations
Methods: Graphical lasso using the top 50 most predictive CpGs, mRNAs, and miRNAs from the final predictive models


A
IQ
SRS
Correlation: −0.47
ASD
Control
Case
B
SRS
P = 4.1 x 10−14
Control
Case
ASD
IQ
P < 2.2 x 10−16
Control
Case
ASD
C
IQ
SRS
Covariate
Smoker
Sex (male)
Race (other)
Race (black)
Placental inflammation
Maternal Age
Insurance: Medicaid
Gestational days
Birthweight Z−score
R2 = 0.12
P = 2.3 x 10-8
R2 = 0.15
P = 2.8 x 10-11
Coefficient estimate


A
IQ
SRS
Mean adjusted R^2
0.00
0.02
0.04
0.06
10^0
10^1
10^2
10^3
10^4
10^5
10^6
Number of features (log_{10} scale)
B
IQ
SRS
Mean adjusted R^2
0.00
0.05
0.10
All
miRNA/CpG
mRNA/CpG
miRNA/mRNA
CpG
miRNA
mRNA
Biomarker type
Biomarker type
CpG
miRNA
mRNA


A
IQ
SRS
P = 6.6 x 10-4
P = 0.03
In−sample predicted value (ELGAN−ECHO)
2
1
0
−1
−2
Control
Case
Control
Case
ASD
ASD
Control
Case
B
IQ
SRS
P = 0.37
P = 0.12
Out−sample predicted value (MARBLES)
2
1
0
−1
−2
Control
Case
Control
Case
ASD
ASD
Control
Case