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

Hudson P Santos Jr., PhD1,2*, Arjun Bhattacharya, PhD3*, Robert M Joseph, PhD4, Lisa Smeester, PhD2,5,6, Karl CK Kuban, MD7, Carmen J Marsit, PhD8, T. Michael O’Shea, MD9, Rebecca C Fry, PhD2,5,6

1 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 Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, 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. Phone: 919.966.9483. Email:

Running title: Neurodevelopmental prediction from placental omics

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

Background: Children born extremely preterm are at heightened risk for intellectual and social impairment. There is increasing evidence for a key role of the placenta in prenatal neurodevelopmental programming, suggesting that the placenta serves as a role in the origins of neurodevelopmental outcomes.

Methods: We examined associations between genomic and epigenomic profiles in the placenta 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. Genome-wide mRNA, CpG methylation and miRNA were assessed 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 then integrated for a predictive analysis of IQ and SRS outcomes using kernel aggregation regression.

Results: We found that genes with important roles in placenta angiogenesis and neural function were associated with intellectual and social impairment. Multi-omic predictions of intellectual and social function were strong, explaining approximately 10% and 12% of the variance in SRS and IQ scores via cross-validation, respectively.

Conclusions: Our findings demonstrate that 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.
Introduction

Despite substantial research efforts to elucidate the etiology of neurodevelopmental impairment (1), little is known about genomic and epigenomic factors influencing trajectories of neurodevelopment. (2) such as those associated with preterm delivery. 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 adolescence 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). 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 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 (10,11). This observation agrees with Boyle et al.’s omnigenic model (12,13), 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, however, has been unexplored.

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 (14). We then combined the genomic and epigenomic data to identify correlative networks of placental genomic and epigenomic biomarkers predictive of social and intellectual impairment. To assess the convergent validity of our behavioral findings, we also examined the association of social and intellectual impairment in relation to ASD diagnoses (15). 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 and Materials**

**ELGAN recruitment and study participants**

From 2002-2004, women who gave birth at <28 weeks gestation at one of 14 medical centers across five U.S. states enrolled in the ELGAN study (14). 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 at 10 years of age**

Tained child psychologist examiner (5,16) evaluated general cognitive ability (IQ) with the School-Age Differential Ability Scales-II (DAS-II) Verbal and Nonverbal Reasoning subscales (17). 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 (18). We used the gender-normed T-score (SRS-T; intended to correct gender differences observed in normative samples) as continuous measure of social deficit (19). All participants were assessed for ASD (15) as described in **Supplementary Methods**.

**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 (20) (see **Supplementary Methods**). 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.

**Epigenome-wide placental DNA methylation**

Extracted DNA sequences were bisulfite-converted using the EZ DNA methylation kit (Zymo Research, Irvine, CA) and followed by quantification using the Infinium MethylationEPIC BeadChip (Illumina, San Diego, CA), which measures CpG loci at a single nucleotide resolution, as previously described (20–23). Quality control and normalization were performed resulting in 856,832 CpG probes from downstream
analysis, with methylation represented as the average methylation level at a single CpG site ($\beta$-value)

Genome-wide placental mRNA and miRNA expression

mRNA expression was determined using the Illumina QuantSeq 3' mRNA-Seq Library Prep Kit, a method with high strand specificity. mRNA-sequencing libraries were pooled and sequenced (single-end 50 bp) on one lane of the Illumina HiSeq 2500. mRNA were quantified through pseudo-alignment with Salmon v.14.0 (27) mapped to the GENCODE Release 31 (GRCh37) reference transcriptome. miRNA expression profiles were assessed using the HTG EdgeSeq miRNA Whole Transcriptome Assay (HTG Molecular Diagnostics, Tucson, AZ). miRNA were aligned to probe sequences and quantified using the HTG EdgeSeq System (28). Genes and miRNAs with less than 5 counts for each sample were filtered, resulting in 11,224 genes and 2,047 miRNAs for downstream analysis. Unwanted technical and biological variation was estimated using RUVSeq (29), where we empirically defined transcripts not associated with outcomes of interest as negative control housekeeping probes (30). One dimension of unwanted variation was removed after upper quantile normalization and adjustments for overdispersion (31–33).

Statistical Analysis

All code and functions used in the statistical analysis can be found at https://github.com/bhattacharya-a-bt/multiomics_ELGAN.

Correlative analyses between SRS, IQ, and ASD

Associations among SRS scores, IQ and ASD were assessed using Pearson correlations with estimated 95% confidence intervals, and the mean difference in SRS and IQ across ASD case-control was assessed using Wilcoxon rank-sum tests. Associations between demographic variables (race, sex, maternal age, number of gestational days, maternal smoking status, and mother’s insurance) with SRS and IQ were assessed using multivariable regression, assessing the significance of regression parameters using Wald tests of significance and adjusting for multiple testing with the Benjamini-Hochberg procedure (34).
Genome-wide molecular associations with SRS and IQ

Once associations between SRS and IQ and ASD were confirmed, we utilized continuous SRS and IQ measures as the outcome. Associations between mRNA expression or miRNA expression with SRS and IQ were estimated through a negative binomial linear model using DESeq2 (33). Epigenome-wide associations (EWAS) of CpG methylation sites with outcomes were assessed using robust linear regression (35) with test statistic modification through an empirical Bayes procedure (36), described previously (23). 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 (34).

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 (31). 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). 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 (37) or kernel regression least squares (KRLS) (38) and assessed predictive performance with McNemar’s adjusted $R^2$ via Monte Carlo cross validation (39). 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 establish external validation. We obtained one external placental CpG methylation dataset from the Markers of Autism Risk in Babies-Learning Early Signs (MARBLES) cohort (40). 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) (40). $\beta$-values for DNA methylation were extracted from BED files and transformed into $M$-values with an offset of 1 (41), 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 (42,43).
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 Supplemental Table 1). Therefore, we considered continuous measures of social impairment (SRS) and cognitive development (IQ) at age 10 for both associative and predictive analyses. 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 ρ = −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 13% of the total variance explained in IQ and SRS, as measured by adjusted $R^2$, with a summary of regression parameters in Supplemental 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.

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 (Supplemental 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 (Supplemental 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).

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 and feature selecting 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 (42,43) (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 (44–49).

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 10% (Spearman $\rho = 0.32, P < 7.6 \times 10^{-14}$) and 12% (Spearman $\rho = 0.35, P < 1.9 \times 10^{-12}$)
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
$-0.56 (W = 8121, P = 6.6 \times 10^{-4})$ for IQ, and mean difference of 0.33 ($W = 4717, P = 0.03$) for SRS
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) (40). 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.
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 10% 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 impairments. For example, CpG methylation in *SLIT3* was associated with intellectual (IQ) disability. *SLIT3* is highly expressed in trophoblastic endothelial cells (50) 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 (48,51–53). *LZTS2*, another gene we found to be associated with IQ, is involved in regulating embryonic development by the Wnt signaling pathway (54,55). Genetic and miRNA expression studies have linked *LZTS2* to social impairment and ASD (56–58). 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 (59–61). 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 (62). In previous studies, *ADAMTS6* has been implicated in intellectual disability and growth development and with socially affected traits in pigs (63,64).

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 (65). 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 (66,67), 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 (68). 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 (69,70). Incorporating information from regulatory biomarkers, like transcription factors and miRNAs, into TWAS increases study power to generate hypotheses about regulation (71,72). 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 (72). These transcriptomic models can then be applied to genome-wide association study cohorts to study the regulation of gene-trait associations in the placenta.

When interpreting the results of this study, some factors should be considered. Extremely preterm birth is strongly associated with increased risk for neurodevelopmental disorders (15). This association may lead to added bias in estimated associations between the molecular biomarkers and outcomes (73). 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, tissue-specific molecular patterns in the placenta should be taken into consideration when interpreting these finding 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.

Our analysis underscores the importance of synthesizing data representing various levels of biological
data to understand distinct genomic and epigenomic 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.
Acknowledgements

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

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).

Data Availability

Multiomic data from the ELGAN study will be made public on the NCBI Gene Expression Omnibus upon publication of the manuscript. Data can be requested from H.P.S. while the manuscript is under review. For validation, we used MethylC-seq data from the MARBLES study available at GSE67615.

Disclosures

The authors have no competing interests to disclose.
1. Hodyl NA, Aboustate N, Bianco-Miotto T, Roberts CT, Clifton VL, Stark MJ (2017, September 1): Child neurodevelopmental outcomes following preterm and term birth: What can the placenta tell us? Placenta, vol. 57. W.B. Saunders Ltd, pp 79–86.

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

3. Agrawal S, Rao SC, Bulsara MK, Patole SK (2018, September 1): Prevalence of autism spectrum disorder in preterm infants: A meta-Analysis. Pediatrics, vol. 142. American Academy of Pediatrics. https://doi.org/10.1542/peds.2018-0134

4. Xie S, Heuvelman H, Magnusson C, Rai D, Lyall K, Newschaffer CJ, et al. (2017): 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 31: 586–594.

5. Korzeniewski SJ, Joseph RM, Kim SH, Allred EN, O’Shea TM, Leviton A, et al. (2017): Social Responsiveness Scale Assessment of the Preterm Behavioral Phenotype in 10-Year-Olds Born Extremely Preterm. J Dev Behav Pediatr 38: 697–705.

6. Rosenfeld CS (2020): The placenta-brain-axis. Journal of Neuroscience Research. John Wiley and Sons Inc. https://doi.org/10.1002/jnr.24603

7. Rosenfeld CS (2020): Placental serotonin signaling, pregnancy outcomes, and regulation of fetal brain development†. Biol Reprod 102: 532–538.

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

9. Paquette AG, Houseman EA, Green BB, Lesseur C, Armstrong DA, Lester B, Marsit CJ (2016): Regions of variable DNA methylation in human placenta associated with newborn neurobehavior. Epigenetics 11: 603–613.

10. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. (2019): Identification of common genetic risk variants for autism spectrum disorder. Nat Genet 51: 431–444.

11. Sullivan PF, Agrawal A, Bulik CM, Andreassen OA, Børglum AD, Breen G, et al. (2018): Psychiatric
Genomics: An Update and an Agenda. *Am J Psychiatry* 175: 15.

12. Boyle EA, Li YI, Pritchard JK (2017): An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 169: 1177–1186.

13. Liu X, Li YI, Pritchard JK (2019): Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* 177: 1022-1034.e6.

14. O'Shea TM, Allred EN, Dammann O, Hirtz D, Kuban KCK, Paneth N, Leviton A (2009): The ELGAN study of the brain and related disorders in extremely low gestational age newborns. *Early Hum Dev* 85: 719–725.

15. Joseph RM, O'Shea TM, Allred EN, Heeren T, Hirtz D, Paneth N, et al. (2017): Prevalence and associated features of autism spectrum disorder in extremely low gestational age newborns at age 10 years. *Autism Res* 10: 224–232.

16. Joseph RM, O'Shea TM, Allred EN, Heeren T, Hirtz D, Jara H, et al. (2016): Neurocognitive and Academic Outcomes at Age 10 Years of Extremely Preterm Newborns. *Pediatrics* 137: e20154343.

17. Beran TN (2007): Elliott, C. D. (2007). Differential Ability Scales (2nd ed.). San Antonio, TX: Harcourt Assessment. *Can J Sch Psychol* 22: 128–132.

18. Constantino JN, Davis SA, Todd RD, Schindler MK, Gross MM, Brophy SL, et al. (2003): Validation of a brief quantitative measure of autistic traits: comparison of the social responsiveness scale with the autism diagnostic interview-revised. *J Autism Dev Disord* 33: 427–433.

19. Constantino JN, Zhang Y, Frazier T, Abbacchi AM, Law P (2010): Sibling Recurrence and the Genetic Epidemiology of Autism. *Am J Psychiatry* 167: 1349–1356.

20. Addo KA, Bulka C, Dhingra R, Santos HP, Jr, Smeester L, et al. (2019): Acetaminophen use during pregnancy and DNA methylation in the placenta of the extremely low gestational age newborn (ELGAN) cohort. *Environ Epigenetics* 5. https://doi.org/10.1093/EEP/DVZ010

21. Bulka CM, Dammann O, Santos HP, VanderVeen DK, Smeester L, Fichorova R, et al. (2019): Placental CpG methylation of inflammation, angiogenic, and neurotrophic genes and retinopathy of prematurity. *Investig Ophthalmol Vis Sci* 60: 2888–2894.

22. Clark J, Martin E, Bulka CM, Smeester L, Santos HP, O'Shea TM, Fry RC (2019): 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 14: 1102–1111.

23. Santos HP, Bhattacharyya A, Martin EM, Addo K, Psioda M, Smeester L, et al. (2019): Epigenome-wide DNA methylation in placentas from preterm infants: association with maternal socioeconomic status. Epigenetics 14: 751–765.

24. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA (2014): Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30: 1363–1369.

25. Fortin J-P, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. (2014): Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol 15: 503.

26. Johnson WE, Li C, Rabinovic A (2007): Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8: 118–127.

27. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017): Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14: 417–419.

28. Qi Z, Wang L, Desai K, Cogswell J, Stern M, Lawson B, et al. (2019): Reliable Gene Expression Profiling from Small and Hematoxylin and Eosin–Stained Clinical Formalin-Fixed, Paraffin-Embedded Specimens Using the HTG EdgeSeq Platform. J Mol Diagnostics 21: 796–807.

29. Risso D, Ngai J, Speed TP, Dudoit S (2014): Normalization of RNA-seq data using factor analysis of control genes or samples. Nat Biotechnol 32: 896–902.

30. Gagnon-Bartsch JA, Speed TP (2012): Using control genes to correct for unwanted variation in microarray data. Biostatistics 13: 539–552.

31. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015): limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43: e47–e47.

32. Bullard JH, Purdom E, Hansen KD, Dudoit S (2010): Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics 11: 94.

33. Love MI, Huber W, Anders S (2014): Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15: 550.
34. Benjamini Y, Hochberg Y (1995): Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple. *Source: Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57. Retrieved August 5, 2019, from https://www.jstor.org/stable/pdf/2346101.pdf?refreqid=excelsior%3A6411207ed4b0feb82c3964cc8b

35. Tsai P-C, Bell JT (2015): Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. *Int J Epidemiol* 44: 1429–1441.

36. Phipson B, Lee S, Majewski IJ, Alexander WS, Smyth GK (2016): ROBUST HYPERPARAMETER ESTIMATION PROTECTS AGAINST HYPERVARIABLE GENES AND IMPROVES POWER TO DETECT DIFFERENTIAL EXPRESSION. *Ann Appl Stat* 10: 946–963.

37. Endelman JB (2011): Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *Plant Genome* 4: 250–255.

38. Hainmueller J, Hazlett C (2014): Kernel Regularized Least Squares: Reducing Misspecification Bias with a Flexible and Interpretable Machine Learning Approach. *Polit Anal* 22: 143–168.

39. Xu QS, Liang YZ (2001): Monte Carlo cross validation. *Chemom Intell Lab Syst* 56: 1–11.

40. Schroeder DI, Schmidt RJ, Crary-Dooley FK, Walker CK, Ozonoff S, Tancredi DJ, et al. (2016): Placental methylome analysis from a prospective autism study. *Mol Autism* 7: 51.

41. Du P, Zhang X, Huang C-C, Jafari N, Kibbe WA, Hou L, Lin SM (2010): Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 11: 587.

42. Friedman J, Hastie T, Tibshirani R (2007): Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 1–10.

43. Epskamp S, Cramer AOJ, Waldorp LJ, Schmittmann VD, Borsboom D (2012): Qgraph: Network visualizations of relationships in psychometric data. *J Stat Softw* 48.

44. Crawley JN, Heyer WD, LaSalle JM (2016, March 1): Autism and Cancer Share Risk Genes, Pathways, and Drug Targets. *Trends in Genetics*, vol. 32. Elsevier Ltd, pp 139–146.

45. Chater-Diehl E, Ejaz R, Cytrynbaum C, Siu MT, Turinsky A, Choufani S, et al. (2019): New insights
into DNA methylation signatures: SMARCA2 variants in Nicolaides-Baraitser syndrome. *BMC Med Genomics* 12: 105.

46. Tang S, Hughes E, Lascelles K, Simpson MA, Pal DK, Marini C, *et al.* (2017): New SMARCA2 mutation in a patient with Nicolaides–Baraitser syndrome and myoclonic astatic epilepsy. *Am J Med Genet Part A* 173: 195–199.

47. Koga M, Ishiguro H, Yazaki S, Horiuchi Y, Arai M, Niizato K, *et al.* (2009): Involvement of SMARCA2/BRM in the SWI/SNF chromatin-remodeling complex in schizophrenia. *Hum Mol Genet* 18: 2483–2494.

48. Glessner JT, Wang K, Sleiman PMA, Zhang H, Kim CE, Flory JH, *et al.* (2010): Duplication of the slit3 locus on 5q35.1 predisposes to major depressive disorder. *PLoS One* 5.

49. Thiffault I, Zuccarelli B, Welsh H, Yuan X, Farrow E, Zellmer L, *et al.* (2017): Hypotonia and intellectual disability without dysmorphic features in a patient with PIGN-related disease. *BMC Med Genet* 18: 124.

50. Liao W-X, Laurent LC, Agent S, Hodges J, Chen D (2012): Human Placental Expression of SLIT/ROBO Signaling Cues: Effects of Preeclampsia and Hypoxia1. *Biol Reprod* 86.

51. Cukier HN, Dueker ND, Slifer SH, Lee JM, Whitehead PL, Lalanne E, *et al.* (2014): Exome sequencing of extended families with autism reveals genes shared across neurodevelopmental and neuropsychiatric disorders. *Mol Autism* 5: 1.

52. Christoforou A, Espeseth T, Davies G, Fernandes CPD, Giddaluru S, Mattheisen M, *et al.* (2014): GWAS-based pathway analysis differentiates between fluid and crystallized intelligence. *Genes, Brain Behav* 13: 663–674.

53. Park SM, Plachez C, Huang S (2018): Sex-dependent motor deficit and increased anxiety-like states in mice lacking autism-associated gene slt3. *Front Behav Neurosci* 12.

54. Stalman SE, Solanky N, Ishida M, Alemán-Charlet C, Abu-Amero S, Alders M, *et al.* (2018): Genetic Analyses in Small-for-Gestational-Age Newborns. *J Clin Endocrinol Metab* 103: 917–925.
55. Thyssen G, Li T-H, Lehmann L, Zhuo M, Sharma M, Sun Z (2006): LZTS2 Is a Novel β-Catenin-Interacting Protein and Regulates the Nuclear Export of β-Catenin. *Mol Cell Biol* 26: 8857–8867.

56. Chang Y, Lin C, Huang H, Chang J, Kuo H (2019): Chromosomal microarray and whole-exome sequence analysis in Taiwanese patients with autism spectrum disorder. *Mol Genet Genomic Med* 7. https://doi.org/10.1002/mgg3.996

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

58. Lackinger M, Sungur AÖ, Daswani R, Soutschek M, Bicker S, Stemmler L, et al. (2019): A placental mammal-specific micro RNA cluster acts as a natural brake for sociability in mice. *EMBO Rep* 20. https://doi.org/10.15252/embr.201846429

59. Gompers AL, Su-Feher L, Ellegood J, Copping NA, Riyadh MA, Stradleigh TW, et al. (2017): Germline Chd8 haploinsufficiency alters brain development in mouse. *Nat Neurosci* 20: 1062–1073.

60. Sugathan A, Biagioli M, Golzio C, Erdin S, Blumenthal I, Manavalan P, et al. (2014): CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. *Proc Natl Acad Sci U S A* 111: E4468–E4477.

61. Yasin H, Gibson WT, Langlois S, Stowe RM, Tsang ES, Lee L, et al. (2019): A distinct neurodevelopmental syndrome with intellectual disability, autism spectrum disorder, characteristic facies, and macrocephaly is caused by defects in CHD8. *J Hum Genet* 64: 271–280.

62. Kelwick R, Desanlis I, Wheeler GN, Edwards DR (2015): The ADAMTS (A Disintegrin and Metalloproteinase with Thrombospondin motifs) family. *Genome Biol* 16: 113.

63. Malli T, Duba HC, Erdel M, Marschon R, Kranewitter W, Deutschbauer S, et al. (2014): Disruption of the ARID1B and ADAMTS6 loci due to a t(5;6)(q12.3;q25.3) in a patient with developmental delay. *Am J Med Genet Part A* 164: 3126–3131.

64. Wu P, Wang K, Yang Q, Zhou J, Chen D, Liu Y, et al. (2019): Whole-genome re-sequencing association study for direct genetic effects and social genetic effects of six growth traits in Large White pigs. *Sci Rep* 9: 1–12.

65. Anastasiadi D, Esteve-Codina A, Pferrer F (2018): Consistent inverse correlation between DNA
methylation of the first intron and gene expression across tissues and species. *Epigenetics* 66. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. (2015): Integrative analysis of 111 reference human epigenomes. *Nature* 518: 317–329.

67. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. (2018): The Encyclopedia of DNA elements (ENCODE): Data portal update. *Nucleic Acids Res* 46: D794–D801.

68. Lea AJ, Vockley CM, Johnston RA, Del Carpio CA, Barreiro LB, Reddy TE, Tung J (2018): Genome-wide quantification of the effects of DNA methylation on human gene regulation. *ELife* 7.

https://doi.org/10.7554/eLife.37513

69. Gamazon ER, Wheeler HE, Shah KP, Mozaffari S V, Aquino-Michaels K, Carroll RJ, et al. (2015): A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* 47: 1091–1098.

70. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, et al. (2016): Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* 48: 245–252.

71. Zhang W, Voloudakis G, Rajagopal VM, Readhead B, Dudley JT, Schadt EE, et al. (2019): Integrative transcriptome imputation reveals tissue-specific and shared biological mechanisms mediating susceptibility to complex traits. *Nat Commun* 10: 3834.

72. Bhattacharya A, Love MI (2020): Multi-omic strategies for transcriptome-wide prediction and association studies. *bioRxiv* 2020.04.17.047225.

73. Paternoster L, Tilling K, Davey Smith G (2017): Genetic epidemiology and Mendelian randomization for informing disease therapeutics: Conceptual and methodological challenges ((G. S. Barsh, editor)). *PLOS Genet* 13: e1006944.
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 (34). 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 and miRNA and EWAS for CpG methylation sites

Clinical associations with IQ and SRS

**Objective:** Find clinical variables associated with outcomes
**Methods:** Multivariable linear regression using ordinary 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 for 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 the 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 model in external datasets
**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 CpG, mRNAs, and miRNAs from final predictive model
