

## Supplementary Methods

### *Projection to kernel space*

For a given omic design matrix  $\mathbf{X}_m$  on  $n$  samples and  $p_m$  features, with each column standard-normalized to zero mean and unit variance, we project the omic into a sparse kernel space using the simple linear kernel:

$$\mathbf{K}_{linear_m} = p_m^{-1} \mathbf{X}_m \mathbf{X}_m^T,$$

resulting in  $\mathbf{K}_{linear_m}$ , an  $n \times n$  matrix. We call  $\mathbf{K}_{linear_m}$  the omic linear kernel matrix for omic  $m$ .

Alternatively, for a given omic  $\mathbf{X}_m$  on  $n$  samples and  $p_m$  features, with each column standard-normalized to zero mean and unit variance, we project the omic design matrix into a sparse Gaussian kernel space using the Gaussian kernel function. For samples  $i$  and  $j$ , corresponding to the  $i$ th and  $j$ th rows of  $\mathbf{X}_m$ , the  $i, j$ -th element of the Gaussian kernel matrix  $\mathbf{K}_{Gauss_m}$  for omic  $m$  is defined as

$$\mathbf{K}_{Gauss_m} = \exp\left(\frac{-\|x_i - x_j\|^2}{p_m}\right).$$

To compare two omic kernels  $\mathbf{K}_{m_1}$  and  $\mathbf{K}_{m_2}$  in the same kernel space (i.e. two linear kernel matrices or two Gaussian kernel matrices), we compute their kernel alignment  $S$ , a measure of similarity between two omic kernels [7] using the standardized Frobenius inner product of two matrices:

$$S = \frac{\langle \mathbf{K}_{m_1}, \mathbf{K}_{m_2} \rangle_F}{\sqrt{\langle \mathbf{K}_{m_1}, \mathbf{K}_{m_1} \rangle_F \langle \mathbf{K}_{m_2}, \mathbf{K}_{m_2} \rangle_F}},$$

where, for two  $n \times n$  matrices  $\mathbf{A}$  and  $\mathbf{B}$ ,  $\langle \mathbf{A}, \mathbf{B} \rangle_F = \sum_{i,j=1}^n A_{ij} B_{ij}$ .

### *Multiomic kernel predictive model with linear kernels*

We consider the following linear model, assuming  $M$  total omics considered:

$$\mathbf{Y} = \mathbf{X}_c \boldsymbol{\beta}_c + \mathbf{U} + \boldsymbol{\epsilon},$$

where  $\mathbf{Y}$  is an outcome of interest (either SRS or IQ), standardized to zero mean and unit variance,  $\mathbf{X}_c$  is the design matrix of covariates,  $\boldsymbol{\beta}_c$  is a vector of fixed effects for covariates,  $\mathbf{U}$  is a vector of omic predictive scores, and  $\boldsymbol{\epsilon}$  is Gaussian random error with zero mean and identity covariance matrix. We assume that

the omic predictive scores are normally distributed with zero mean and a fused kernel matrix as its variance-covariance matrix, i.e.  $\mathbf{U} \sim N(0, \mathbf{K})$  and  $\mathbf{K} = \sum_{m=1}^M \mathbf{K}_m$ . This model is simply a traditional linear mixed model, treating clinical covariates as fixed effects and aggregating all omics into omic predictive scores that are treated as random effects clustered around zero. It is straightforward that

$$\mathbf{U} = \sum_{m=1}^M \mathbf{U}_m = \sum_{m=1}^M \sum_{j=1}^{p_m} \mathbf{X}_{mj} \eta_{mj},$$

where  $\mathbf{U}_m$  is the vector of omic predictive scores for the  $m$ th omic profile and the weight sum of omic features  $\mathbf{X}_{mj}$  with random effects coefficients  $\eta_{mj}$ . Assuming  $\eta_{mj} \sim N(0, \sigma_m^2/p_m)$ , we see that  $\mathbf{U}_m \sim N(0, \sigma_m^2 \mathbf{K}_m)$  and  $\mathbf{U} \sim N(0, \mathbf{K})$ .

#### *Kernel modeling training and evaluation for linear kernels*

We evaluate our models through 50-fold Monte Carlo cross-validation, using methods similar to Zhu et al [8]. Given a single cross-validation fold, with 75% of the data in training set and 25% in a test set, we can

divide  $\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_{train} \\ \mathbf{Y}_{test} \end{bmatrix}$  and  $\mathbf{K}_{linear}$  into four blocks:

$$\mathbf{K}_{linear} = \begin{bmatrix} \mathbf{K}_{train} & \mathbf{K}_{cov} \\ \mathbf{K}_{cov}^T & \mathbf{K}_{test} \end{bmatrix},$$

where  $\mathbf{K}_{train}$  is the variance matrix of the training set,  $\mathbf{K}_{test}$  is the variance matrix of the test set, and  $\mathbf{K}_{cov}$  is the covariance matrix between the training and test sets. We find the maximum likelihood estimator  $\hat{\boldsymbol{\beta}}_C$  for the fixed effects, the best linear unbiased predictor  $\hat{\mathbf{U}}_{train}$  of the omic predictive scores of the training set, and the restricted maximum likelihood estimators  $\hat{\sigma}_m^2$  for the variance components of the fused kernel, using *rrBLUP* [9]. We then estimate the total predictive scores  $\hat{\mathbf{Y}}_{test}$ :

$$\hat{\mathbf{Y}}_{test} = \mathbf{X}_{C_{test}} \hat{\boldsymbol{\beta}}_C + \mathbf{K}_{cov} \mathbf{K}_{train}^{-1} \hat{\mathbf{U}}_{train}$$

We then compute the adjusted  $R^2$  between the  $\mathbf{Y}_{test}$  and  $\hat{\mathbf{Y}}_{test}$  to assess the predictive performance of the omic predictive scores. This process is repeated across 100 folds and the adjusted  $R^2$  are averaged to create a predictive index for a given clinical and omic profile. We consider all possible combinations of omics after regressing out clinical covariates in these multiomic kernel models. External validation is

conducted similarly; however, as the relevant clinical covariates were not available in the MARBLES dataset, we do not adjust the molecular profiles for these clinical covariates in the predictive models applied to this dataset[10].

#### *Kernel regularized least squares regression for Gaussian kernels and model evaluation*

We employ a similar Monte Carlo cross-validation scheme, splitting our data into 75%-25% training and test sets across 50 folds. We implement kernel-based regularized squares regression through the *KRLS* package [11], that minimizes the Tikhonov objective function over squared loss. Briefly, the objective of KRLS, in our case, is to find the  $c$  that minimizes

$$T(c) = \sum_{i=1}^n (\mathbf{Y} - \mathbf{K}_{Gauss_m} c)^T (\mathbf{Y} - \mathbf{K}_{Gauss_m} c) + \lambda c^T \mathbf{K}_{Gauss_m} c,$$

where  $\lambda$  is tuned via leave-one-out cross-validation. Furthermore, KRLS can compute the pointwise partial derivatives of the fitted function with respect to each predictor using estimators developed by Hainmuller and Hazlett [11]. These pointwise partial derivatives can be used to examine the marginal effect of every feature in the omic design matrix on the outcome of interest, similar to an ordinary least squares regression parameter estimate.

Using the best parameter estimates from the training set, a Gaussian kernel matrix can be computed from the test set, and we define predicted values in the test set as

$$\hat{\mathbf{Y}}_{test} = \mathbf{K}_{Gauss_{m_{test}}} \hat{c}.$$

We compute adjusted  $R^2$  between the observed and predicted values of the outcome of interest here, as well, to assess predictive performance.

#### *Sparse regression model and evaluation*

We also consider linear models to predict outcomes of interest using a regularized regression model:

$$\mathbf{Y} = \mathbf{X}_m \boldsymbol{\beta}_m + \boldsymbol{\epsilon},$$

where  $\mathbf{X}_m$  and  $\boldsymbol{\beta}_m$  are the  $n \times p_m$  design matrix and effects for a given omic profile  $m$ , using Monte Carlo cross-validation to evaluate predictive performance across 100 80%-20% training-test set folds. In the training set, using *glmnet*, we estimate  $\hat{\boldsymbol{\beta}}_m$  using elastic net regularized regression with a mixing parameter

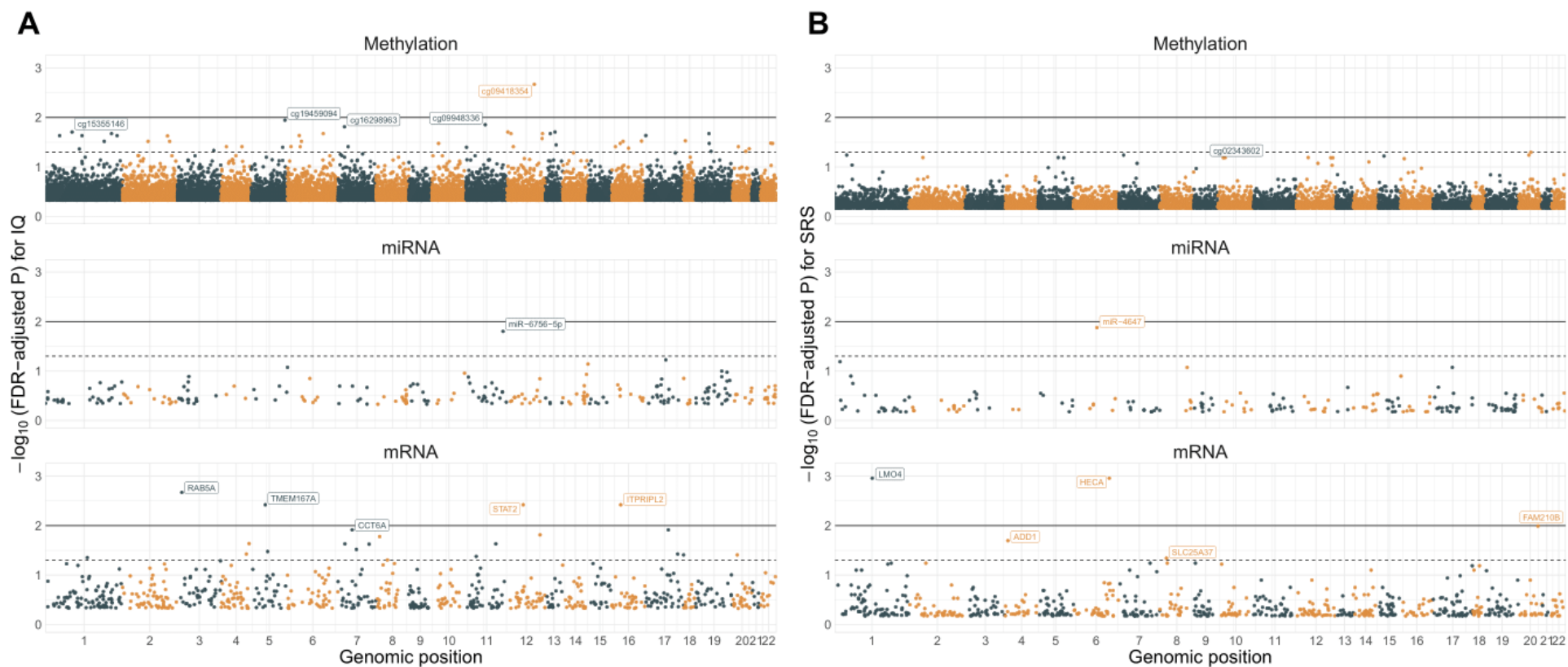
of 0.5 (even mixture of LASSO and ridge penalties) and the regularization penalty parameter tuned over 5 folds [12]. Using these parameter estimates, we predict on the test set and evaluate predictive performance using adjusted  $R^2$ .

#### *Feature selection*

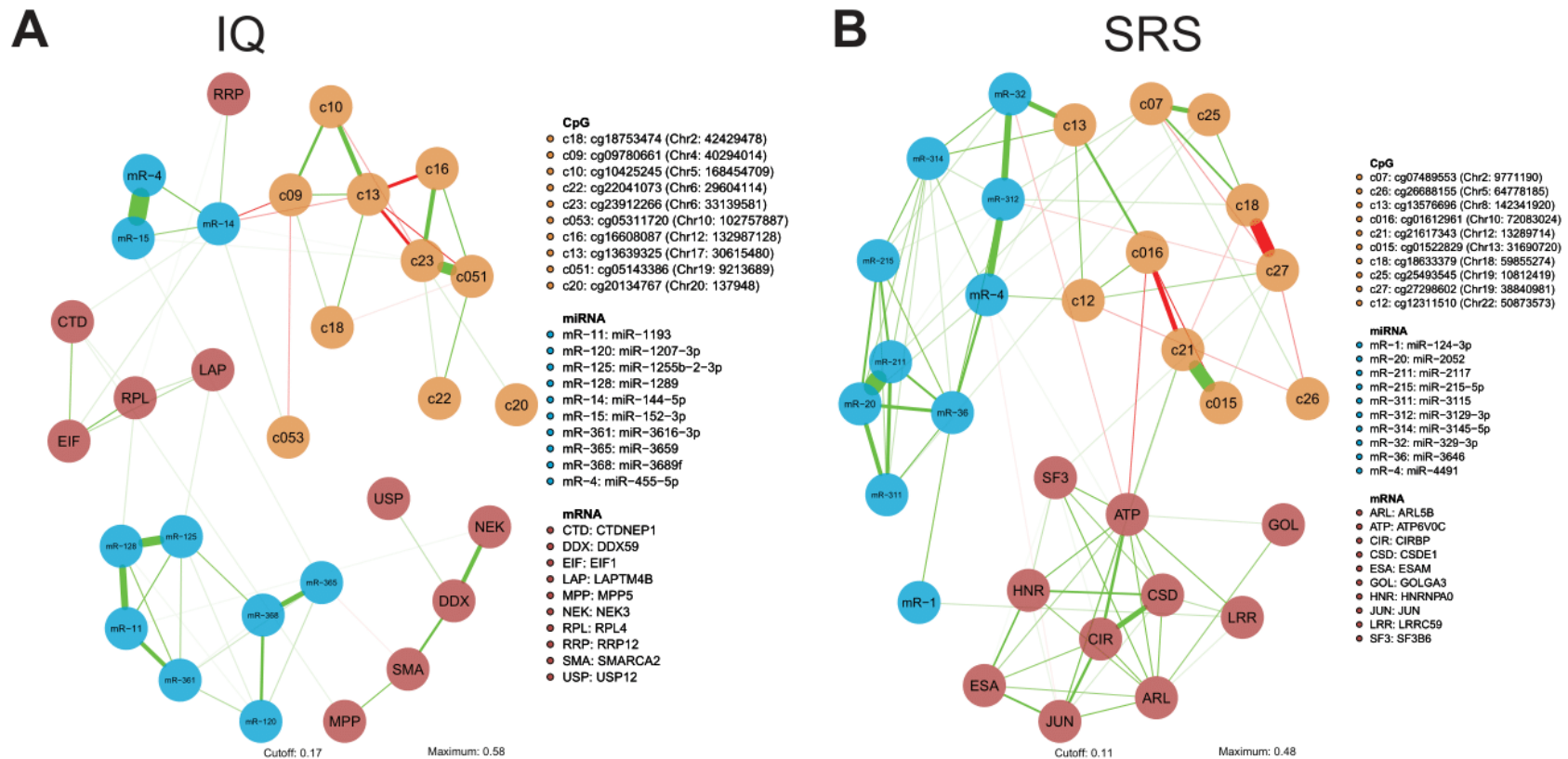
To tune the number of features to include for each omic in the final predictive models, we tuned over various  $P$ -value thresholds from fold-wise one-way tests of associations. We split the data into 5 80%-20% training set-test set folds, conducted one way tests of associations in the training set (i.e. differential expression analysis for mRNA and miRNA expression, EWAS for DNA methylation), and selected all biomarkers with associations with the outcome of interest with  $P$ -value under a given threshold. Using these biomarkers, we predict the outcome of interest in the test set using both kernel regression methods and compute the adjusted  $R^2$  to assess predictive performance.

#### *External validation using MARBLES dataset*

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]. We extracted  $\beta$ -values for DNA methylation from BED files and transformed into  $M$ -values with an offset of 1 [13]. We then used the best linear kernel and kernel regression models to predict SRS and IQ in the MARBLES dataset, as detailed above. It is important to note that not all CpG sites used in the best-methylation model from ELGAN were assayed in the MARBLES external validation set (only approximately 85% overlap). Furthermore, the MARBLES dataset does not have measures of SRS or IQ. Thus, to assess the validity of predicted SRS and IQ estimates in MARBLES, we tested for association between the predicted SRS and IQ values and ASD case-control status.



*Supplemental Figure 1: **Manhattan plots of one-way omic association tests for IQ and SRS.*** Manhattan plots for one-way tests of association for methylation (top), miRNA (middle), and mRNA expression with IQ (A) and SRS (B). The *X*-axis plots the genomic position of the biomarker and the *Y*-axis plots the Benjamini-Hochberg FDR-adjusted *P*-value for the association with the given outcome. The dotted line provides a reference of FDR-adjusted *P* = 0.05, and the solid line provides a reference of FDR-adjusted *P* = 0.01. Biomarkers are labelled with their association has FDR-adjusted *P* = 0.01.



**Figure 2: Sparse correlative networks between top predictive biomarkers of IQ and SRS.** Using the top 50 CpGs (orange nodes), top 50 miRNAs (blue), and top 50 mRNAs (red) that are predictive of IQ (left) and SRS (right), we inferred sparse correlative networks [14, 15]. Nodes are biomarkers and edges show correlations between biomarkers. Positive correlations are shown in green, negative correlations are shown in red, and the thickness of the edge gives the absolute magnitude of correlation.

## References

1. Rutter, Bailey, Lord. The social communication questionnaire. Western Ps. Los Angeles; 2003.
2. Joseph RM, O'Shea TM, Allred EN, Heeren T, Hirtz D, Paneth N, et al. Prevalence and associated features of autism spectrum disorder in extremely low gestational age newborns at age 10 years. *Autism Res.* 2017;10:224–232.
3. Risi S, Lord C, Gotham K, Corsello C, Chrysler C, Szatmari P, et al. Combining information from multiple sources in the diagnosis of autism spectrum disorders. *J Am Acad Child Adolesc Psychiatry.* 2006;45:1094–1103.
4. Lord C, Rutter M, DiLavore P, Risi S, Gotham K, Bishop S. *Autism Diagnostic Observation Schedule.* Los Angeles: Western Psychological Corporation; 2012.
5. Joseph RM, O'Shea TM, Allred EN, Heeren T, Hirtz D, Jara H, et al. Neurocognitive and Academic Outcomes at Age 10 Years of Extremely Preterm Newborns. *Pediatrics.* 2016;137:e20154343.
6. Santos HP, Bhattacharya A, Martin EM, Addo K, Psioda M, Smeester L, et al. Epigenome-wide DNA methylation in placentas from preterm infants: association with maternal socioeconomic status. *Epigenetics.* 2019;14:751–765.
7. Elisseeff A, Technologies B, Shawe-Taylor J, Holloway R, Kandola J. On Kernel-Target Alignment N ello Cristianini BIOwulf Technologies nello@support-vector. net.
8. Zhu B, Song N, Shen R, Arora A, Machiela MJ, Song L, et al. Integrating Clinical and Multiple Omics Data for Prognostic Assessment across Human Cancers. *Sci Rep.* 2017;7:16954.
9. Endelman JB. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *Plant Genome.* 2011;4:250–255.
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.* 2016;7:51.
11. Hainmueller J, Hazlett C. Kernel Regularized Least Squares: Reducing Misspecification Bias with

- a Flexible and Interpretable Machine Learning Approach. *Polit Anal.* 2014;22:143–168.
12. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw.* 2010;33:1–22.
  13. 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.* 2010;11:587.
  14. Epskamp S, Cramer AOJ, Waldorp LJ, Schmittmann VD, Borsboom D. Qgraph: Network visualizations of relationships in psychometric data. *J Stat Softw.* 2012;48.
  15. Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics.* 2007:1–10.