Supplemental methods

Surrogate variable analysis (SVA)

The SVA with the full sample resulted in the identification of 15 surrogate variables (SVs), with the first 2 SVs representing 27 and 20% of the variance respectively and the others less than 6% each (Figure S2A). We applied linear regression to each SV using each known covariate as potential predictor. We tested 8 known covariates: Run (2 levels = dates when 450k arrays were processed), Chip (18 levels = Chip carrying the sample), location (96 levels = sample location on the array in each run), Plate_column (8 levels = column in 96-well plates), Plate_row (12 levels = row in 96-well plates), Gender (2 levels), Race (24 levels = self-declared ethnicity), and Age (continuous). For the discrete variables with more than two levels, the multiple regression model was built with dummy variables. This results in models that include different amounts of variables. To compare the different models against each other, we obtained the R squared values, adjusted for the number of variables in each model (Figure S2B). Most surrogate variables are best explained by technical covariates and less by biological covariates. However, it is clear that a lot of the variance remains unexplained by known covariates and justifies our use of the SVs in our model to better account for unknown undesirable variation.

Sensitivity analysis for family effect

The data includes 44 sets of siblings/cousins. To investigate whether this impacts our results, we ran an additional differential methylation analysis on the full sample, using the same model that includes the clinical factor (FASD vs. Control) + all 15 SVs identified, with the addition of 'Family' as a random effect, using the lme4 R package. We compared the results from this analysis to the original analysis. The Spearman rank correlation is 0.9996. There are 1658 significant probes, 1601 overlap with the 1661 probes from the original analysis, 639 overlap with the 658 final probes after ethnicity

adjustment (Table S1). The difference is minor and we conclude that the presence of families in our cohort do not impact significantly our results.

PAE vs FASD analysis

We performed an additional differential methylation analysis after excluding 27 FASD samples with no official diagnosis (denoted PAE for Prenatal Alcohol Exposure). This analysis, including 96 controls and 83 FASDd (FASD with diagnosis), resulted in 502 significant DM probes at a FDR of 0.05, 461 of which overlap with the 1661 probes from the original analysis. The Spearman rank correlation between the results from the original analysis and the analysis without PAE is 0.7927. In addition, we ran ANOVA analyses for the 658 significant probes after ethnicity adjustment, on Control vs. FASDd, Control vs. PAE, and FASDd vs. PAE. We observed that, for most of the probes, methylation in controls is significantly different from either FASDd or PAE, while FASDd vs. PAE doesn't lead to any significant results (Figure S9). We concluded from these results that the PAE samples are not significantly different from the FASDd samples and should be kept in the analysis as part of the broad FASD group.

Investigating ethnically-biased probes

To investigate ethnically-biased probes (denoted "ethnic" probes below), we ran a differential methylation analysis on the FASD samples, looking for differences between samples from the 2 ethnic clusters identified in the MDS analysis (49 samples in cluster 1 vs. 53 samples in cluster 2). It is important to note that the ethnicity bias is directional – the FASD Group is confounded with the ethnic cluster 2 and this is only problematic when the direction of change is the same (e.g. higher in FASD and higher in cluster 2). We thus calculated the one-sided p values and rank for all probes in the above analysis and investigated separately the up-methylated and down-methylated "ethnic" probes. We obtained 1105 probes significantly differentially up-methylated (FDR 0.05) and 594 probes

differentially down-methylated (FDR 0.05) in the ethnic cluster 2. These overlap with 5 up-methylated and 4 down-methylated probes in FASD respectively, out of the 658 significant probes after ethnicity adjustment (Table S1). To investigate whether the ethnicity adjustment we performed was efficient at removing "ethnic" probes, we looked at the "ethnic" ranking of our significant FASD DM probes (Figure S4A). We performed a Receiver Operating Characteristic (ROC) analysis with the ethnic probes (Figure S4B). Our approach was to question how good the ethnic ranking was at predicting the 658 FASD DM probes, a high AUC indicating that the DM probes are in the top of the "ethnic" rank and thus very highly biased. Our results show that while the "ethnic" ranking can somewhat predict our 1661 FASD DM probes (AUC = 0.819 and 0.773 for up- and down-methylated probes respectively), it is much less effective at predicting our 658 FASD DM probes after ethnicity adjustment (AUC = 0.656 and 0.594 for up- and down-methylated probes respectively). These results confirm that our ethnicity adjustment is reducing the ethnic bias as expected.

Incorporating ethnicity as a set of covariates in the regression model

We investigated the option of incorporating the ethnicity as a covariate in the model, using the genotyping data available. There are 195 samples with both genotyping and methylation data, 103 FASD and 92 controls. We tested different models and decided to perform a linear regression analysis using Limma with a model that includes the clinical factor (FASD vs. Control) + Gender + C1 + C2 + C3 (First 3 components from the MDS analysis) + family as a random factor. The multiple-testing adjusted significance threshold was established at $p = 1.2 \times 10^{-7}$ (0.05/404030). No significant probes could be identified. This is expected because the FASD status is confounded with ethnicity and correcting for one will erase the small changes the can be observed in the other.