Supplemental figures

Fig S1. Principal component analysis of buccal epithelial cell and blood samples. The first two principal components, which typically associate with cell type, are depicted for the buccal epithelial cell (BEC) samples from the present study and blood samples obtained from the Gene Expression Omnibus (those beginning with GSM). As shown here, both tissue types cluster separately and little to no blood contamination of BECs is apparent in the dataset.

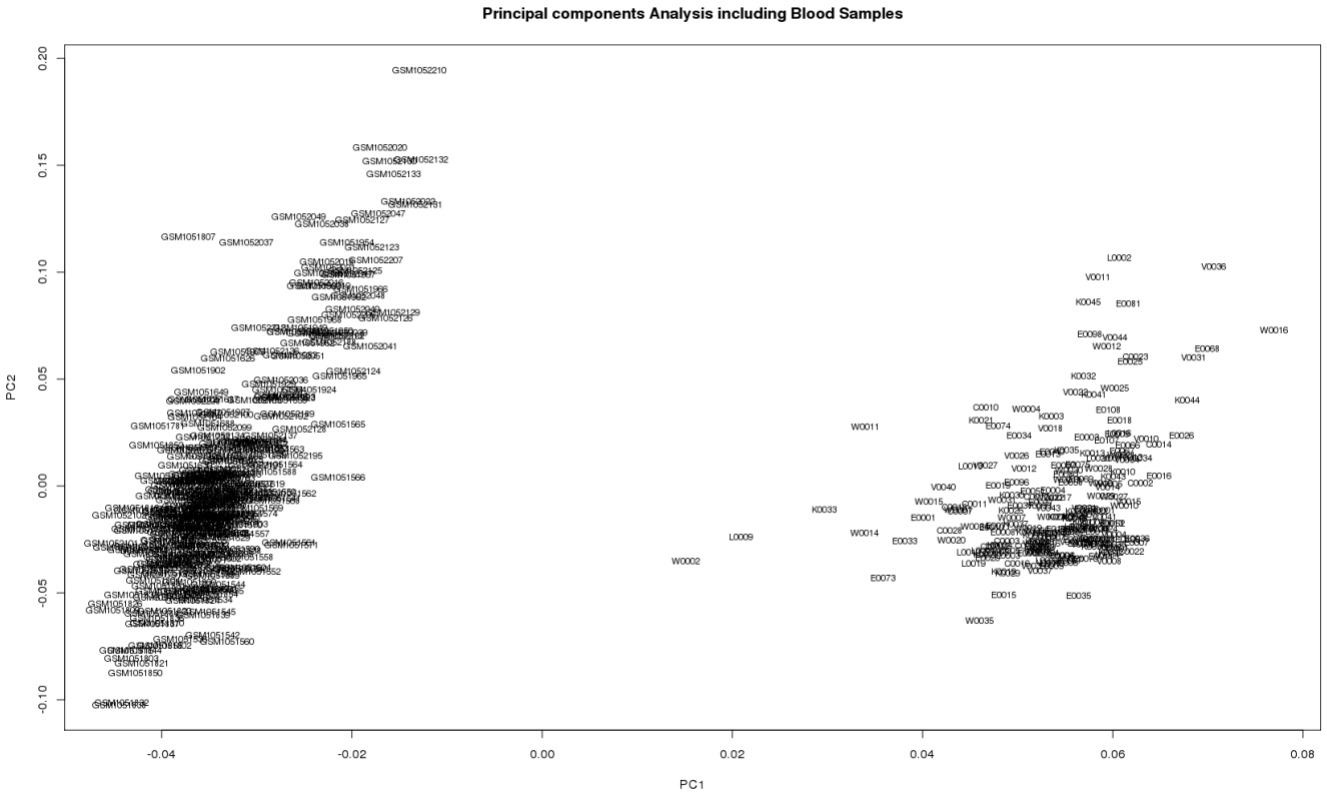
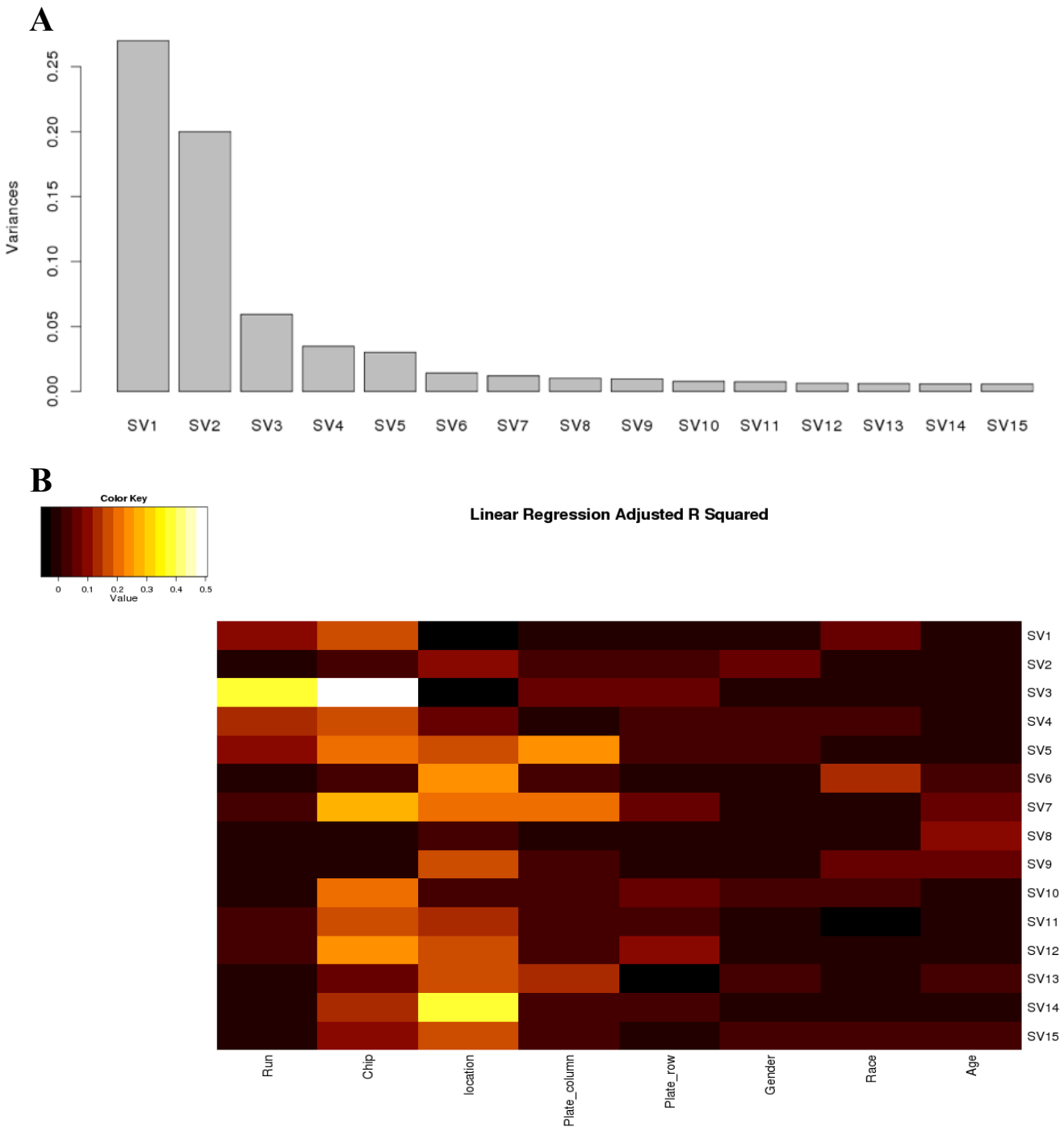


Fig S2. Exploration of surrogate variables. A. Scree plot of the percentage of variance represented by each surrogate variable generated from the entire dataset. B. Heatmap of R squared values from linear regression modeling of each SV with each known covariate. C. Heatmap of adjusted p-values from linear regression modeling of each SV with known covariates and genetic clusters (C1 to C4).



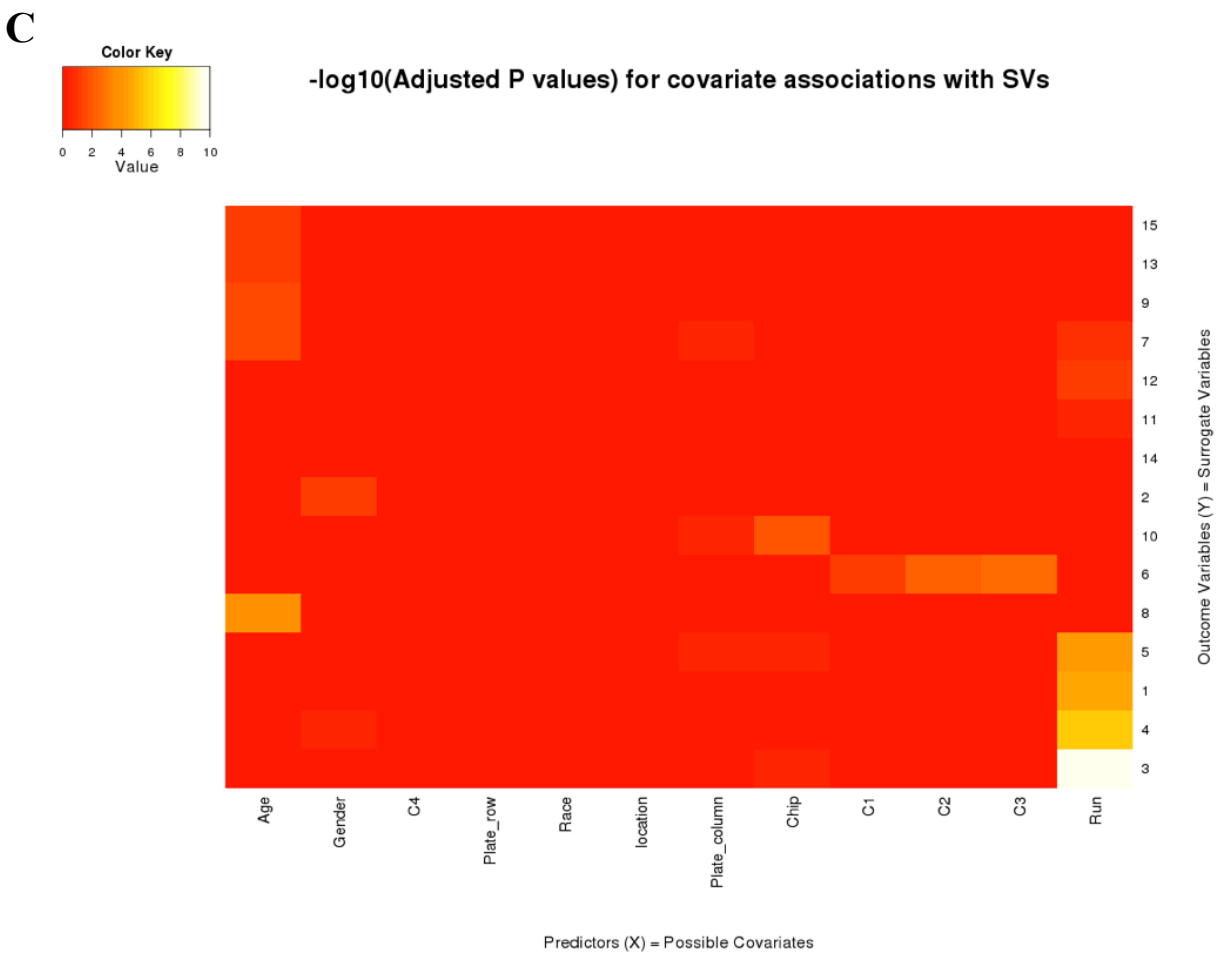


Fig S3. Hierarchical clustering of individual based on MDS analysis of genotyping data. Children with FASD (red) and controls (black) were clustered based on genetic scores from SNP genotyping. Two major subgroups were identified from the unsupervised hierarchical clustering.

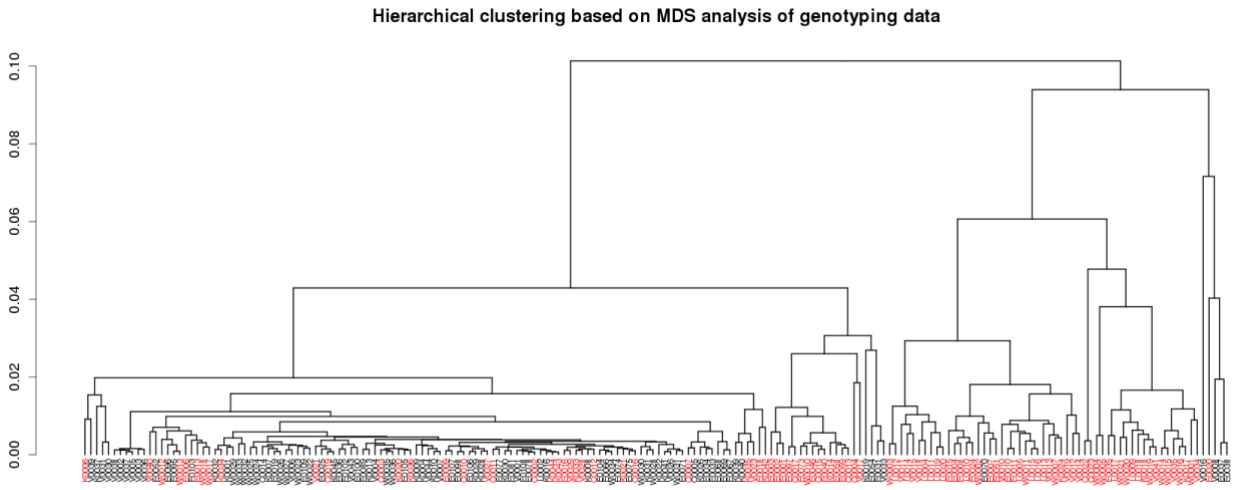
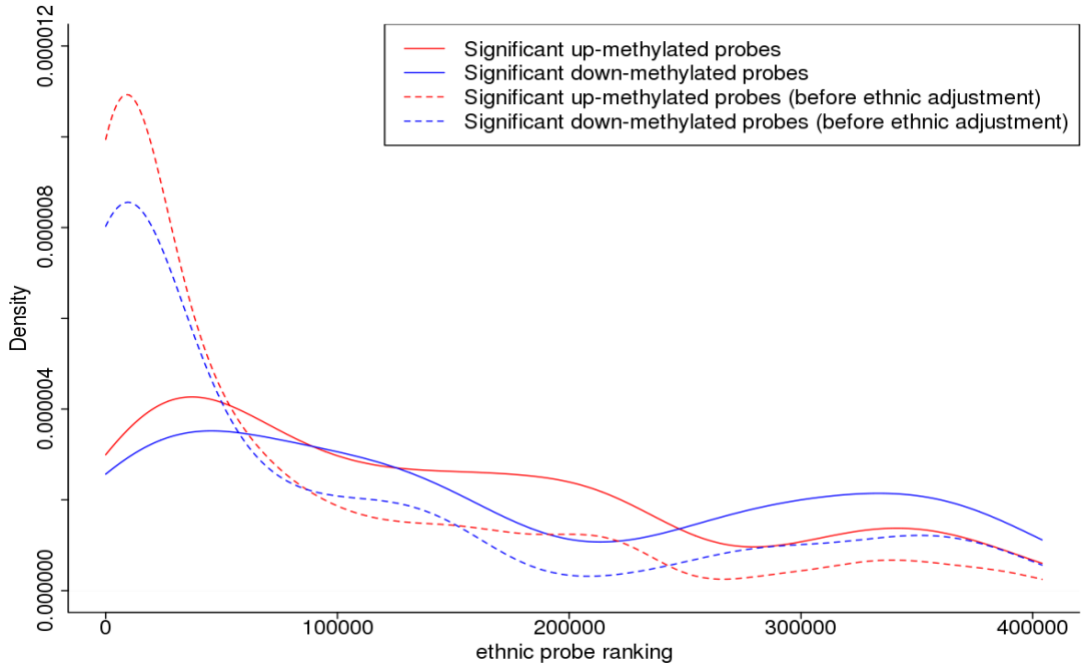


Fig S4. Analysis of ethnically-biased probes in the 658 DM probes. A. Distribution of "ethnic" rank in the significantly up- and down-methylated probes before and after ethnicity adjustment. B. ROC curve demonstrating that the "ethnic" probes are less able to predict the FASD DM probes after ethnicity adjustment.

A



B

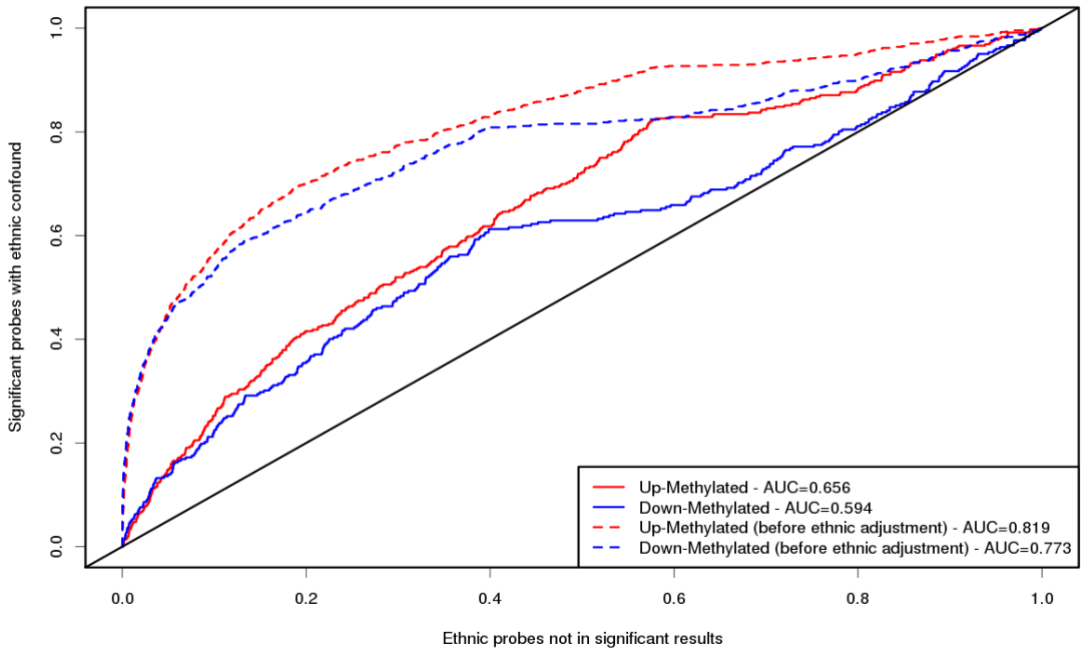


Fig S5. Distribution of socio-economic status scores for children in the FASD and control groups. A. In the full dataset, the distribution of socio-economic status (SES) scores is significantly different between groups (p=0.00017). B. By contrast, in the more ethnically-homogeneous subgroup, the difference is no longer significant (p=0.16), suggesting that these effects may have been partially corrected through the method used for genetic background correction.

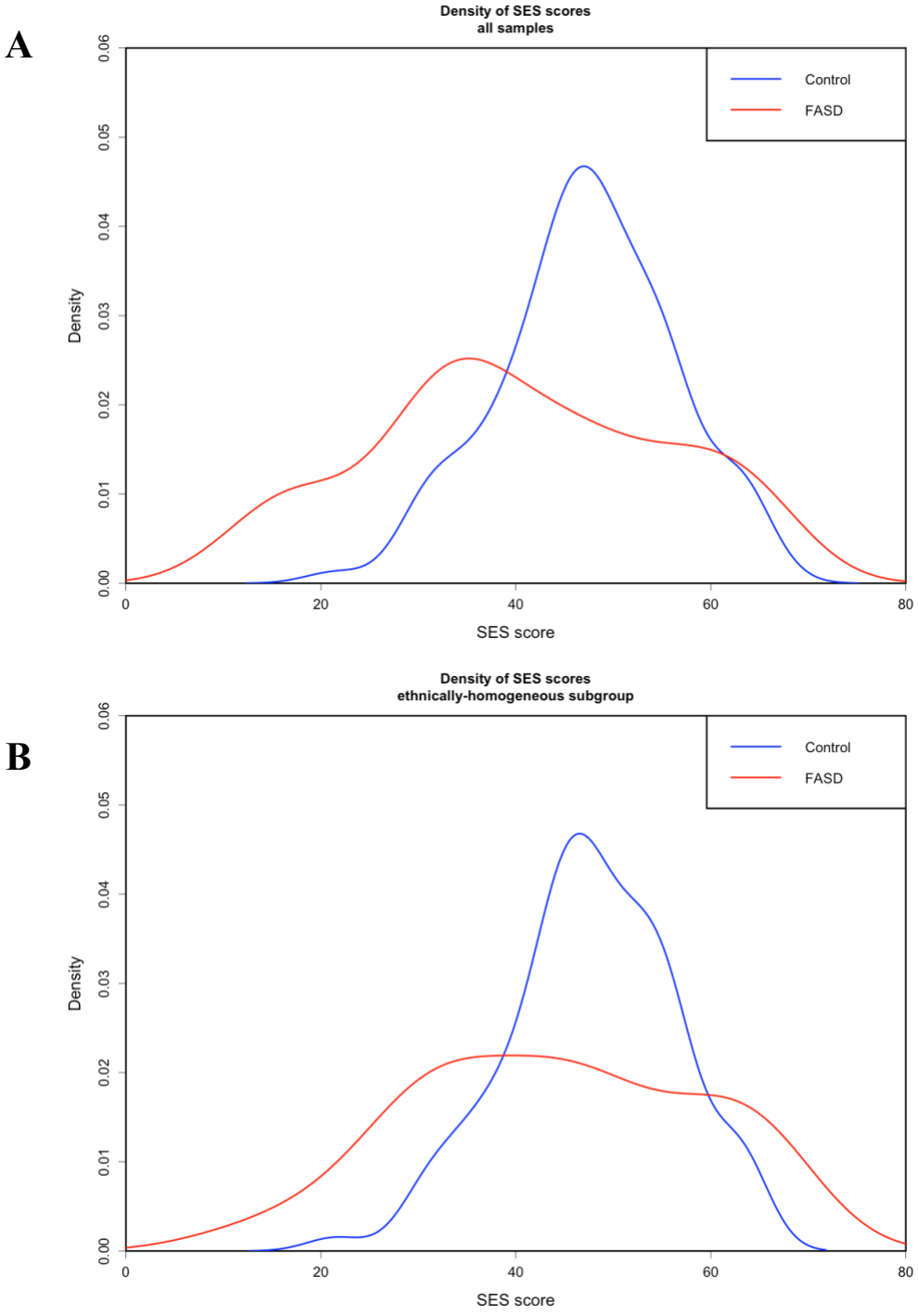


Fig S6. Pyrosequencing verification. Top row are Bland-Altman plots with the difference between the methylation values from the array and the pyrosequencing on the Y axis, and the average Beta values on the X axis. Bottom row are plots with the methylation level from the array on the Y axis and from the pyrosequencing on the X axis, with linear regression lines drawn as well as the Pearson correlation coefficient r.

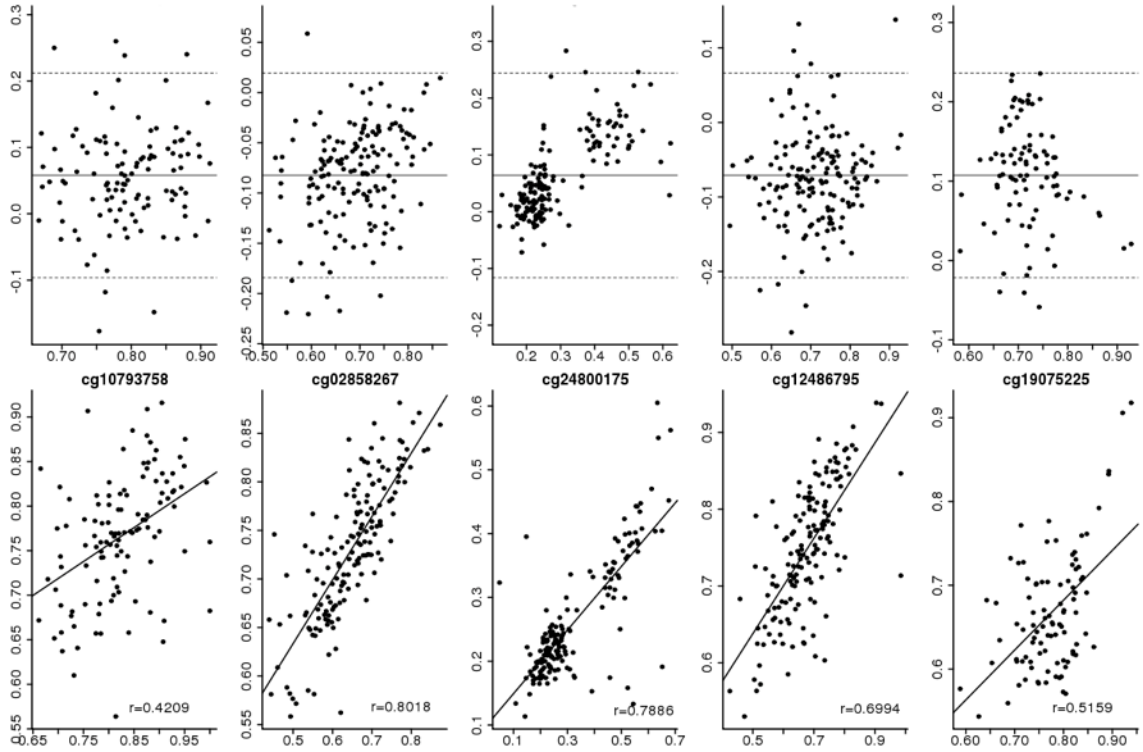


Fig S7. Correlation of buccal and brain methylation at 658 DM CpGs. Points represent an individual CpG methylation in brain and buccal. The overall correlation of mean methylation between buccal and brain samples was 0.76.

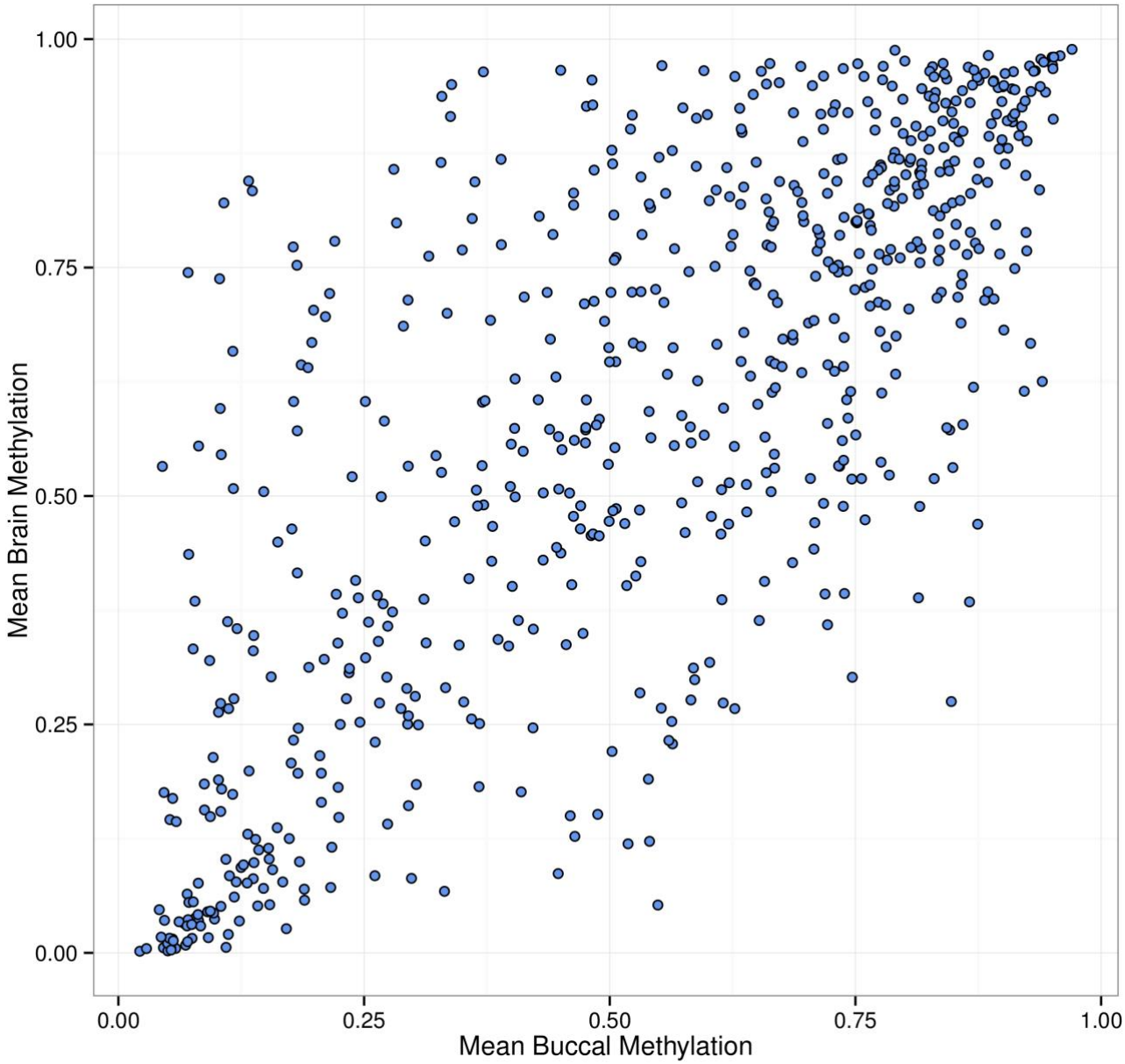


Fig S8. Average distribution of DNA Methylation. The distribution of average DNA methylation for all probes on the 450K array displayed a bimodal distribution, with probes clustering in the lowly (0-20%) and highly (80-100%) methylated ranges. By contrast, probes that were significantly differentially methylated in children with FASD versus controls were more prevalent in intermediately methylated regions (Student's t test; p = 2.5e-09).

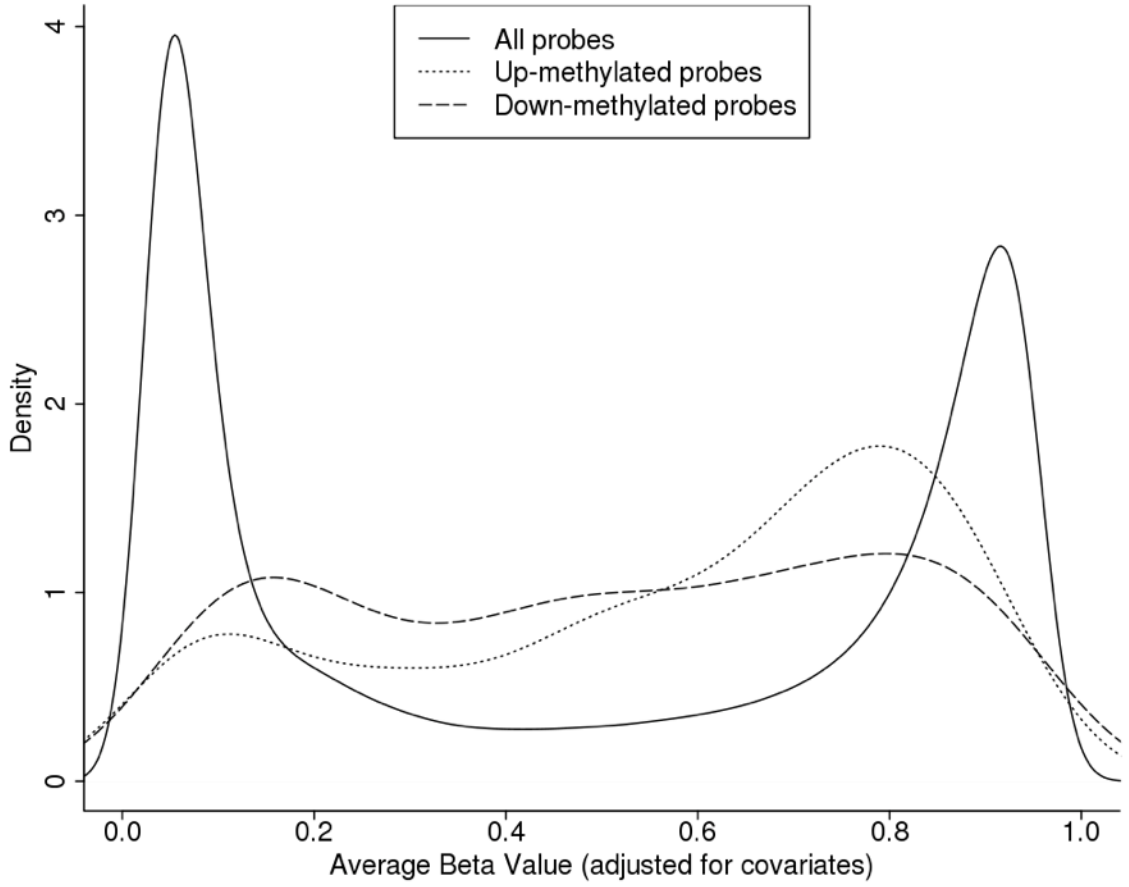


Fig S9. P-value distributions from ANOVA analyses on 658 significant probes. P-value distributions obtained from linear regression analyses were skewed to the left in controls versus both diagnosed FASD cases (FASDd) and children with PAE, but undiagnosed for FASD. By contrast, the p-value distribution comparing both groups of children with PAE revealed a relatively flat distribution, suggesting that there little differences in DNA methylation patterns between diagnosed and undiagnosed children with PAE.