Supplementary Information for
The Biological Embedding of Early Life Socioeconomic Status
and Family Adversity in Childrens Genome-wide DNA
Methylation

Supplementary Text

Additional Adjustments for Effects of Race/Ethnicity

To account for additional potential confounds of race with tested associations, parental-report of childs Ethnic Minority Group status was also included in models. Due to the small N for most subgroups and moderate percentage of multiethnic participants, Ethnic Minority Group status was calculated as a dichotomous variable reflecting minority versus Caucasian groups.

Blood Cell Contamination

In order to control for potential confounding of a variable of interest and contamination of our buccal samples with blood cells, and more specifically T cells, we used in silico spike in of blood and T cell methylation samples to test for contamination [37]. We spiked in 44 T cell samples (GSE50222 and GSE53191) and 37 whole blood (GSE41169 and GSE52113), age matched as best as possible within the limitations of data available. Using PCA the first two PCs associated with tissue type (Supplementary Figure S1).

Genetic Ancestry Additional Control

As an additional control for effect of genetic ancestry on methylation, beyond including genetic ancestry and ethnic minority as covariates, any CpGs associated with genetic ancestry were filtered from differential methylation hit lists. In total 8,445 CpGs associated with genetic ancestry (FDR <0.25, delta beta >0.05, ANOVA). Despite including genetic ancestry as a covariate in all models, there were still a substantial number of CpGs associated with genetic cluster in the differential methylation hit lists (percent of differentially methylated CpGs associated with genetic ancestry 27%-41%).

Sensitivity Analysis

We performed a sensitivity analysis to test whether our results were sensitive to the inclusion of gender and blood cell contamination as both factors would introduce substantial noise into the data. In addition to the covariates included in original models (Table S4) gender and blood contamination were included as covariates. Blood contamination was represented by PC2 from the blood cell contamination PCA as PC2 related most to samples skewing toward the cluster of spiked-in blood samples. At all CpGs which were significantly differentially methylated from the original models (FDR<0.2; delta beta>0.05) for Income-per-Dependent, Parental Education and Family Adversity were tested with models including all covariates (genetically-determined ancestry, self-reported ethnic minority status, child age, twin status plus child gender and PC2 to represent blood contamination). These new nominal p values and delta betas were then correlated with the values from the original models (Figure S3).

Exploration of Adjacent 450K CpGs

As CpGs adjacent to one another are often seen to correlate in DNAm, we explored CpGs adjacent to our significant EWAS CpGs. We looked at CpGs within 1kb of any EWAS CpGs for each variable. We then looked at the delta beta of these adjacent CpGs and compared it to the direction and magnitude of the delta beta from the significant EWAS CpGs (Figure S5).

Bisulfite Pyrosequencing Verification

Three CpGs identified as significantly differentially methylated were chosen to verify the 450K array with Bisulfite PCR-pyrosequencing. The CpGs were chosen had the highest delta betas seen with each variable (Family adversity: cg10581375, Income-per-dependent: cg21502834, Parental Education: cg26511075). Bisulfite PCR-pyrosequencing assays were designed with PyroMark Assay Design 2.0 (Qiagen, Valencia, CA, USA). Regions containing the CpG targets were amplified by PCR using HotstarTaq DNA polymerase kit (Qiagen, Valencia, CA, USA). The quantitative levels of methylation for each CpG were calculated with Pyro Q-CpG software (Qiagen, Valencia, CA, USA). Verification of the 450K was measured as the Spearman correlation and root mean square error between the 450K and pyrosequencing methylation values at the three target CpGs.

CpG to Gene Associations

There are multiple approaches for associating a CpG to a gene, such as the closest TSS [58] or whether the CpG is present in a genes body or promoter [62]. Here we have used a CpG to gene association definition that allows for a CpG to be associated with multiple gene features, as well as multiple genes [63]. This inclusive association is an attempt to capture all possible roles of a CpG in gene regulation. The 485,512 CpGs on the 450K array associated with 23,018 genes (43.8% intragenic CpGs, 34.2% promoter CpGs, 2.5% 3 region CpGs, and 19.5% intergenic CpGs).

Genomic Feature Enrichment

Enrichment in CpG resort features (CpG islands, shores and shelves) and genomic regions was done using significantly differentially methylated CpG lists at FDR<0.2 and delta beta >0.05. Then 1,000 random lists of CpGs were taken as the background, each significant CpG list. The count of CpGs in each resort feature and gene region were used to build a fold change compared to the background and permutation p value, then corrected for multiple comparisons using Benjamini-Hochberg correction [38].
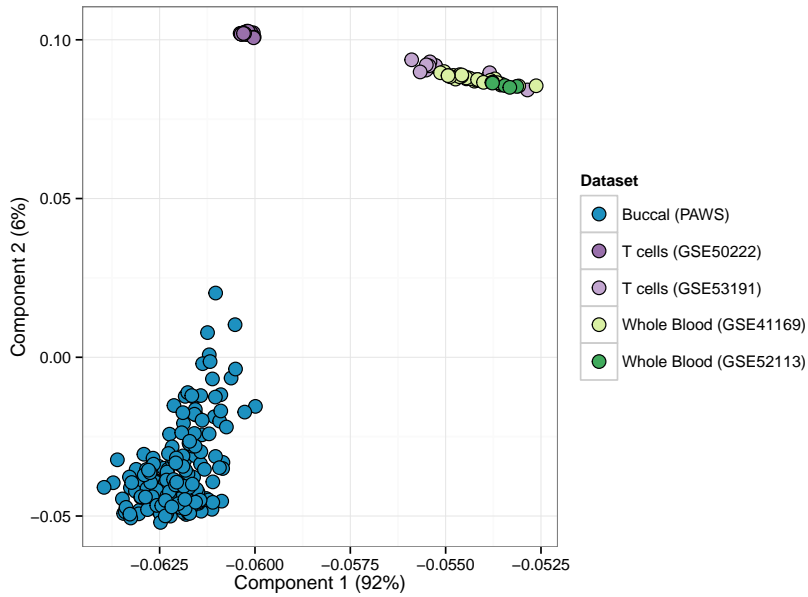
Supplementary Figures

Figure S 1: 4 U V E Z samples show variable levels of blood and t cell contamination. Shown are the first two PCs in the methylation data with blood and t cell tissues spiked in.
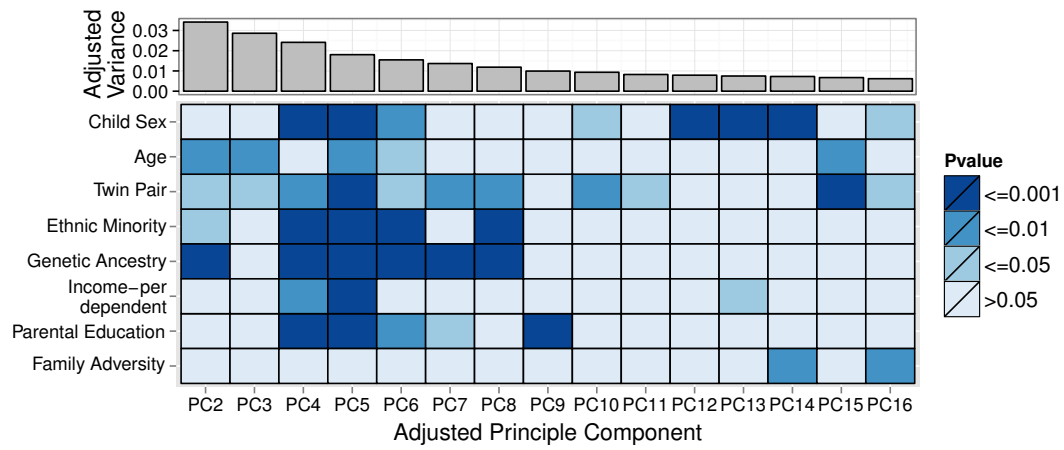

Figure S 2: PCA of meta data variables to assess batch effects. The scree plot shows the amount of methylation variance (adjusted for PC0) accounted for by each PC. Heat map shows the association (correlation p value for continuous; ANONA p value for categorical) between a meta data variable and an individual PC, before ComBat After ComBat, PCA was rerun and the scree plot and heat map show the association with the new PCs.
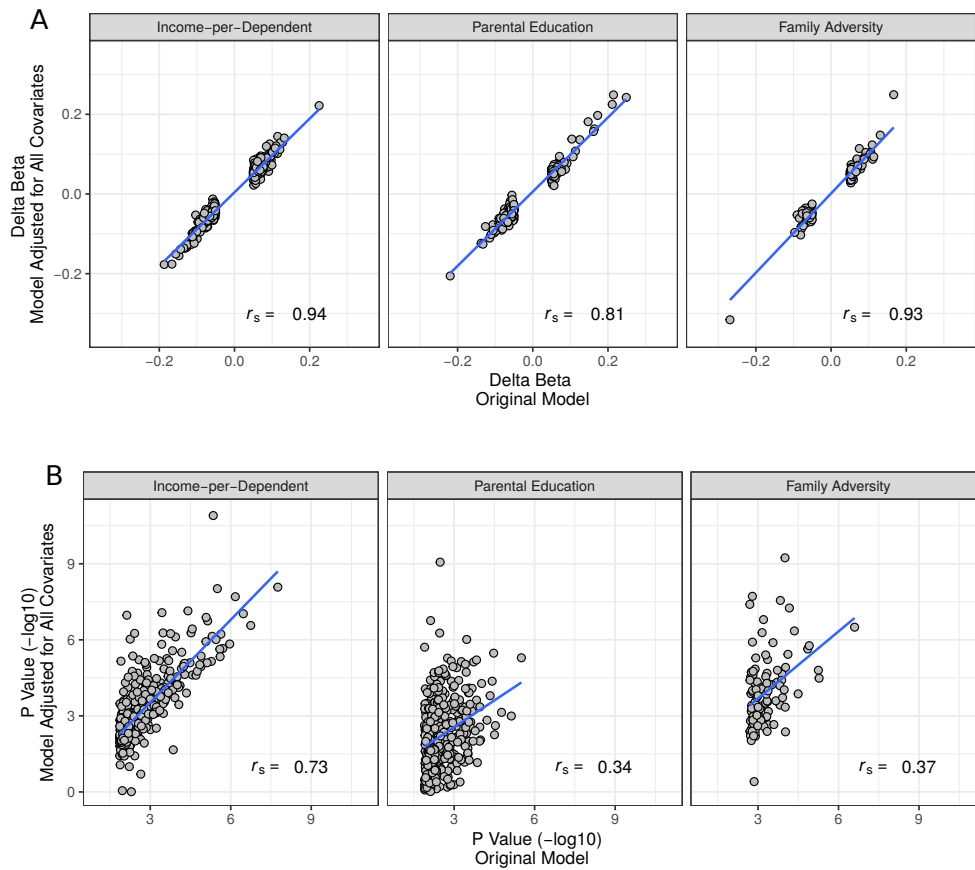
5

Figure S 3: Trends in differential DNAm are maintained irrespective of additional covariate inclusion. CpGs which had observed differential DNAm (FDR<0.2; delta beta>0.05) in original model are shown for each main effect variable. From both the original minimal models and models including gender and blood cell composition plots show the correlation between A) Delta betas B) Nominal p values.
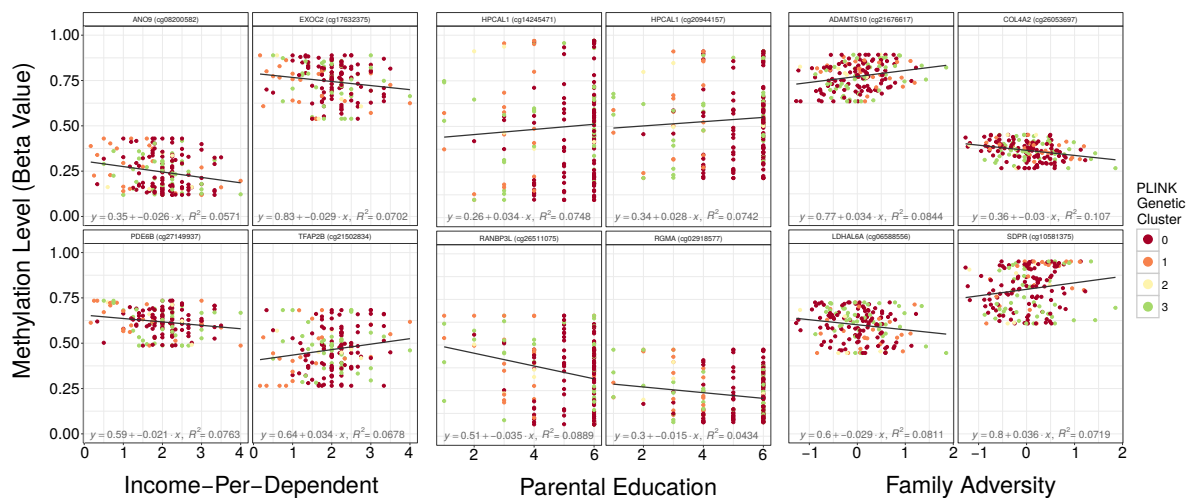
Figure S 4: Trends in differential DNAm are maintained even after 90% winsorization of DNAm values. Plots show the relationship between 90% winsorized DNAm and the variable of interest at representative CpGs (same as those presented in Figure 2C). Each plot is labeled with the CpG ID and the associated gene. Lines show a linear model fit through the data. See also Figures S1-S4 and S6.

7
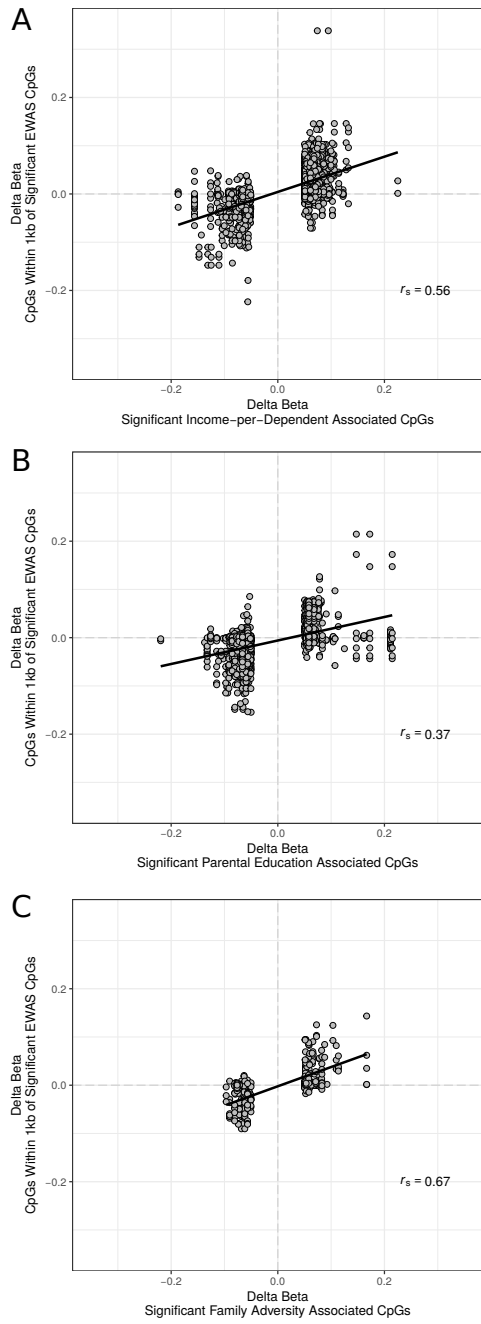
Figure S 5: Trends in DNAm delta beta are consistent between adjacent CpGs within 1kb. Significant EWAS CpGs are plotted against the delta betas of all CpGs within 1kb. Delta betas and EWAS CpGs are shown for each variable A) Income-per-dependent B) Parental Education C)Family Adversity.
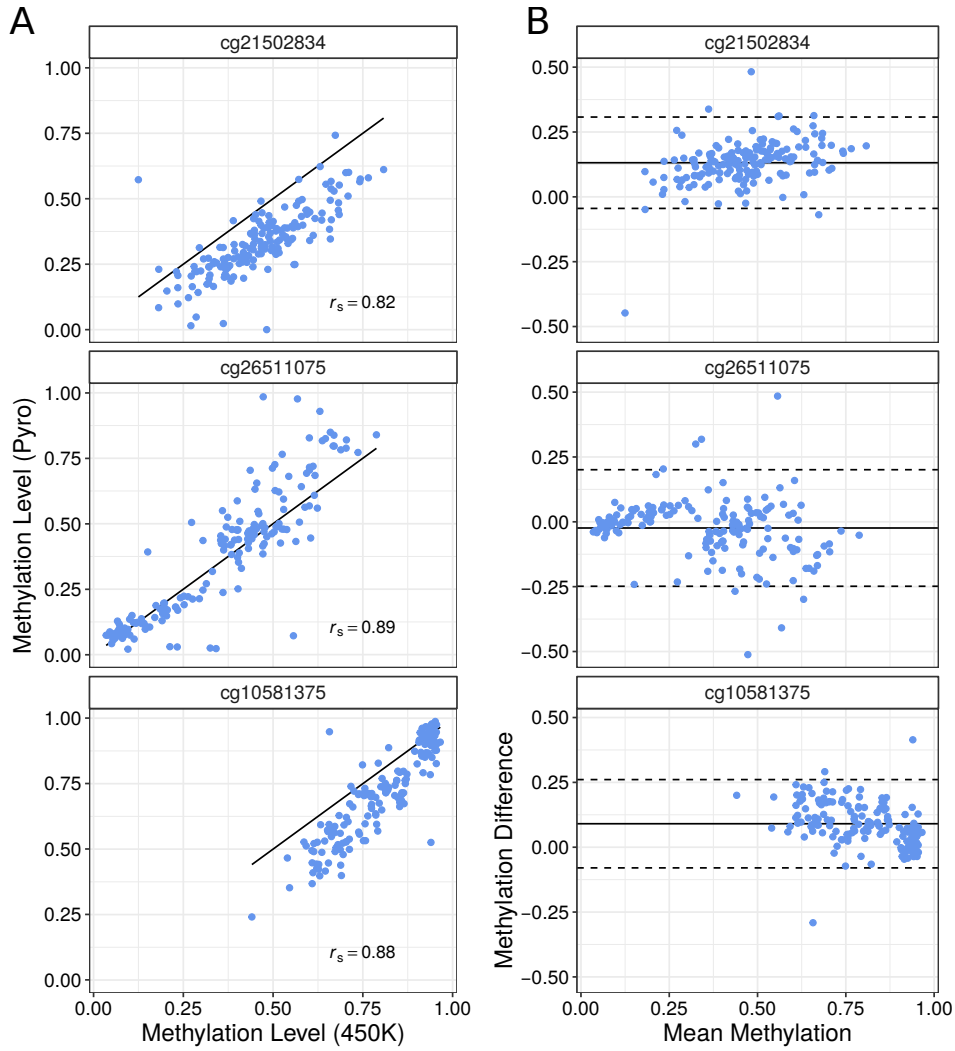
Figure S 6: Bisulfite pyrosequencing verification of the methylation levels measured on the 450K. A) Correlation between 450k methylation and pyrosequencing methylation measures. B)Bland altman plots show that there is no systematic bias of either measurement platform toward any methylation level.
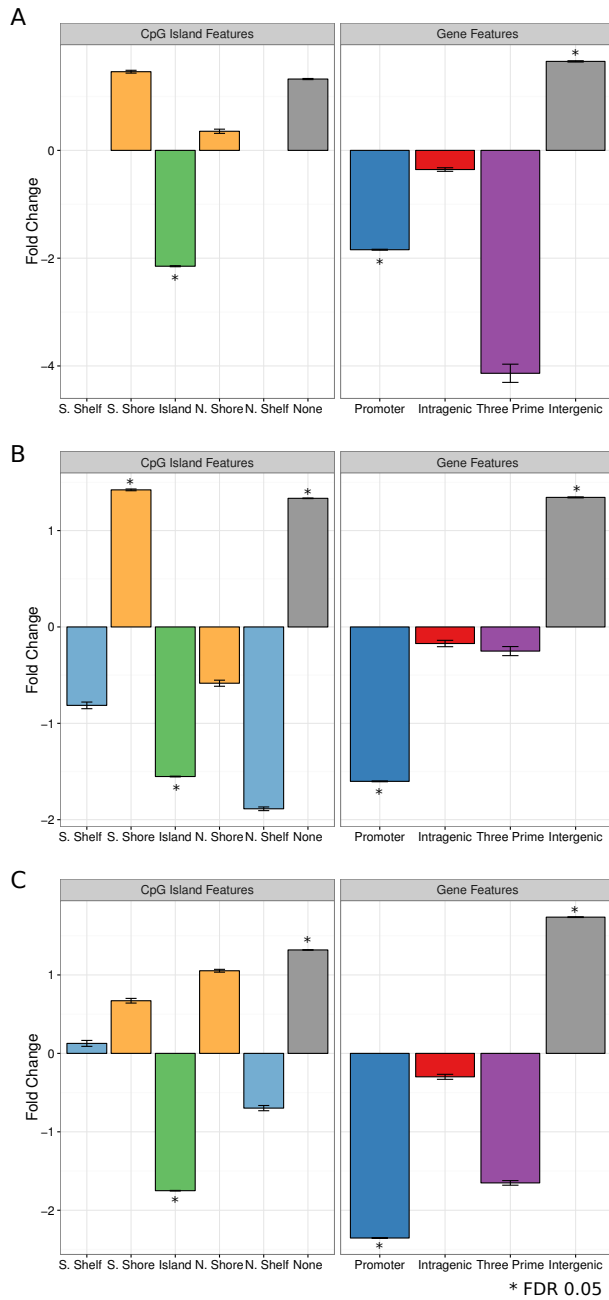
Figure S 7: Differentially methylated CpGs associated with each variable localize to specific genomic features. In all plots bars show the fold change between significantly differentially methylated CpG count in each region and count of CpGs from 1,000 permutations of random CpGs. Error bars show standard error. A) Genomic enrichment is shown for CpGs at which methylation is significantly associated with A) Family adversity B) Parental education C) Income-per-dependent

Figure S 8: Replicate pairs improved in RMSE overall after preprocessing steps.

Supplementary Tables

Table S 1: Demographics of the full cohort and the sub sample with genotyping data collected.

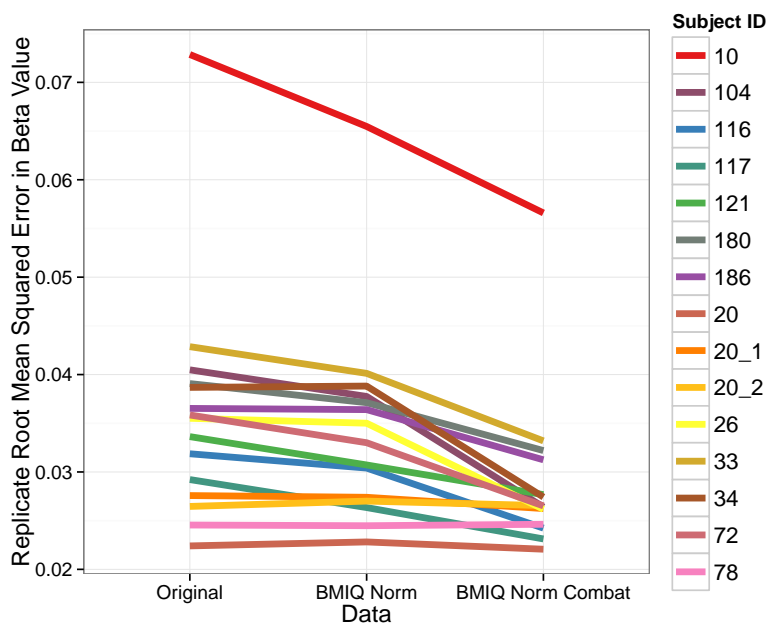| Variable	Original Sample (N=338) N(%)	Mean	SD	Genotyped Sample (N=192) N(%)	Mean	SD
Age at Kindergarten Entry		5.32	0.32		5.34	0.31
Age at Anthropometric Assessment		9.99	0.47		10.03	0.47
Sex						
Female	163(48.2%)			96 (50%)		
Male	175 (51.8%)			96 (50%)		
Race/Ethnicity						
Caucasian	137(43%)			101(54.3%)		
African American	60(19%)			20(10.9%)		
Asian	34(11%)			16(8.7%)		
Latino	13(4%)			6(3.3%)		
Multiethnic	70(22%)			37(19.2%)		
Other	6(1.5%)			4(2.2%)		
Income		7.12			7.67	2.36
less than $10,000 (1)	13(4.1%)			6(3.2%)		
10−19,999 (2)	16(5.1%)			4(2.1%)		
20−29,999 (3)	16(5.1%)			9(4.8%)		
$30-319,999 (4)	22(7.0%)			2(1.1%)		
40−49,999 (5)	15(4.7%)			9(4.8%)		
50−59,999 (6)	17(5.4%)			9(4.8%)		
60−79,999 (7)	40(12.7%)			27(14.3%)		
$80-99,999 (8)	50(15.8%)			36(19.0%)		
100−149,999 (9)	74(23.4%)			52(27.5%)		
150−199,999 (10)	39(12.3%)			28(14.8%)		
more than $200,000 (11)	14(4.4%)			7(3.7%)		
Education		4.68	1.44		4.99	1.34
less than high school degree (1)	8(2.5%)			5(2.6%)		
completed high school (2)	18(5.6%)			6(3.1%)		
some college or 2-yr degree (3)	55(17.1%)			18(9.4%)		
four year college degree (4)	57(17.7%)			31(16.2%)		
some graduate/professional school (5)	39(9.6%)			27(14.1%)		
professional or graduate degree (6)	145(45%)			104(54.5%)		

Table S 2: Family Adversity Subscale Composite Descriptives and Reliabilities for Genetic Subsample

Variable	Scale	Min	Max	M	SD	Alpha
Overall Adversity Composite (average of 6 standardized indices)		-1.27	1.85	0.036	0.56	
Financial Stress	1-5	1	5	2.34	0.97	0.81
Parenting Overload	1-5	1.2	5	3.22	0.74	0.80
Marital Conflict	1-5	1.1	3.1	1.76	0.42	0.74
Parental Depression	1-4	1	2.9	1.41	0.33	0.83
Harsh/Restrictive Parenting	1-7	1.78	6.56	3.55	0.71	0.79
Anger Expression Composite (average of 2 standardized measures)	N/A	-1.94	2.55	0.06	0.85	N/A
Negative Expression	1-9	1.55	7.2	4.15	1.08	*
Anger Expression	0-6	0.63	6	2.43	0.83	*

* Negative Expression Scale was created by combining Negative Subdominant, alpha=.77, and Negative Dominant, alpha=.85. Anger Expression scale was created by combining 3 subscale scores (Anger Out, alpha=.65, Anger In, alpha=.73, and Expression Control, alpha=.74)

Table S 3: Correlations among Adversity Composite Indices

	Financial Stress	Parenting Overload	Marital Conflict	Parental Depression	Harsh/Restrictive Parenting
Financial Stress					
Parenting Overload	0.09				
Marital Conflict	0.15^t	0.12			
Parental Depression	0.14	0.34***	0.17*		
Harsh/Restrictive Parenting	0.19**	0.01	0.05	0.15^t	
Anger Expression	0.06	0.31***	0.39***	0.31***	0.16^t

* p<0.05
** p<0.01
*** p<0.001
^{t} p <0.1

Table S 4: Models and covariates used to find methylation associations.

Main Effect	Covariates	Genetic Ancestry Association P value (ANOVA)	Ethnic Minority Report Association P value (T Test)
Income-per-dependent	Genetic Ancestry, Ethnic Minority, Twin Status, Age	0.0045	0.0152
Parental Education	Genetic Ancestry, Ethnic Minority, Twin Status	<0.0001	<0.0001
Family Adversity	Genetic Ancestry, Twin Status	0.1237	0.2576

Table S 5: CpGs associated with income-per-dependent, family adversity and education. Columns provided are: Gene associated with the CpG, Number of CpG associated with the variable also associated the gene, Number of CpG on the 450K associated with the gene, Enrichment of the gene for variable associated CpGs, Surprise is a metric of how surprising it is to see that number of hits in the gene dependent on the number of CpGs in the gene (CpG associated / Enrichment fromAverage), CpG ID, Chromosome genome build 37, Coordinate genome build 37, Region of the gene the CpG is associated with, Gene isoform, CpG island name if applicable, Relation to the CpG island, Association p value, and the delta beta measure of change in methylation with the variable of interest.

Table S 6: Top over-represented GO groups in the genes with differential methylation associated with each of the three predictors. Columns are: name of the GO gene set, GO ID, nominal p value (which the data is sorted by), Benjamini-Hochberg corrected p value and Multifunctionality (MF) scores

Name	ID	Pval	CorrectedPvalue	MFPvalue
Income-per-dependent				
cell-cell adhesion	GO:0098609	0.00001262	0.08393772	0.922
cell adhesion	GO:0007155	0.00001965	0.06536008	0.918
biological adhesion	GO:0022610	0.00002075	0.04599579	0.918
forelimb morphogenesis	GO:0035136	0.0001177	0.19564612	0.62
positive T cell selection	GO:0043368	0.0001308	0.17399375	0.702
hindlimb morphogenesis	GO:0035137	0.0001678	0.18603108	0.77
positive regulation of developmental process	GO:0051094	0.0001704	0.16187079	0.997
regulation of developmental process	GO:0050793	0.0001882	0.15647559	0.984
single organismal cell-cell adhesion	GO:0016337	0.0002473	0.18271872	0.964
regulation of multicellular organismal process	GO:0051239	0.0002565	0.17061645	0.996
Parental Education				
positive regulation of cell differentiation	GO:0045597	0.0002228	1	0.994
positive regulation of biomineral tissue development	GO:0070169	0.0002674	0.88928181	0.785
positive regulation of developmental process	GO:0051094	0.0008444	1	0.997
positive regulation of osteoblast differentiation	GO:0045669	0.001762	1	0.836
response to carbohydrate	GO:0009743	0.002064	1	0.906
positive regulation of neuron projection development	GO:0010976	0.002064	1	0.879
skeletal system morphogenesis	GO:0048705	0.002188	1	0.969
regulation of activin receptor signaling pathway	GO:0032925	0.002989	1	0.627
negative regulation of retinoic acid receptor	GO:0048387	0.002989	1	0.056
regulation of neuron differentiation	GO:0045664	0.004137	1	0.979
Family Adversity				
negative T cell selection	GO:0043383	0.00009208	0.61243261	0.583
negative thymic T cell selection	GO:0045060	0.00009208	0.61243261	0.582
thymic T cell selection	GO:0045061	0.0001287	0.42804206	0.572
T cell selection	GO:0045058	0.0005518	1	0.783
cellular monovalent inorganic cation homeostasis	GO:0030004	0.00126	1	0.799
T cell differentiation in thymus	GO:0033077	0.002086	1	0.818
thymocyte aggregation	GO:0071594	0.002086	1	0.818
negative regulation of proteolysis	GO:0045861	0.00363	1	0.94
positive regulation of GTPase activity	GO:0043547	0.003693	1	0.339
monovalent inorganic cation homeostasis	GO:0055067	0.004323	1	0.848