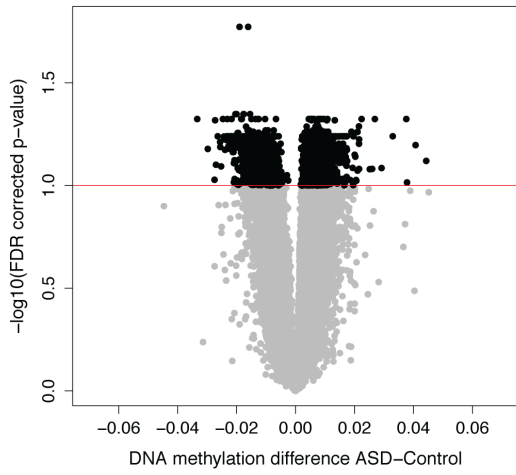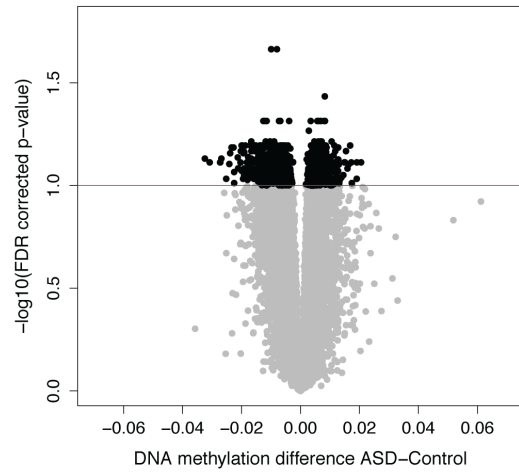Integrative genomics identifies a convergent molecular subtype that links epigenomic with transcriptomic differences in autism

Ramaswami et al.
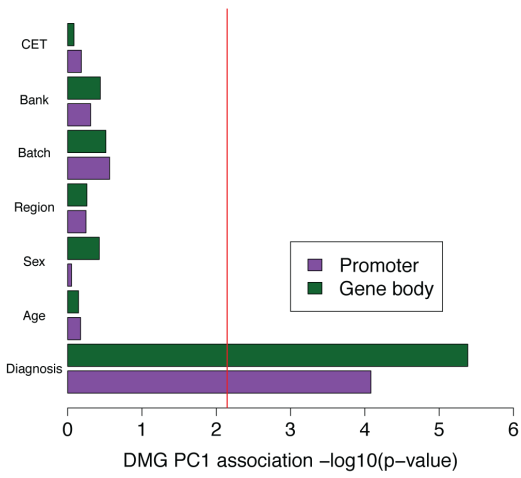
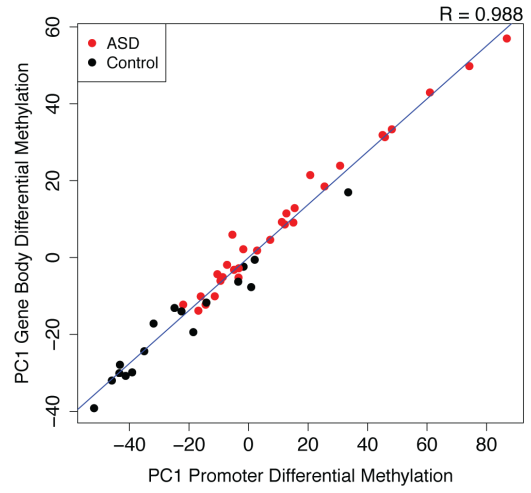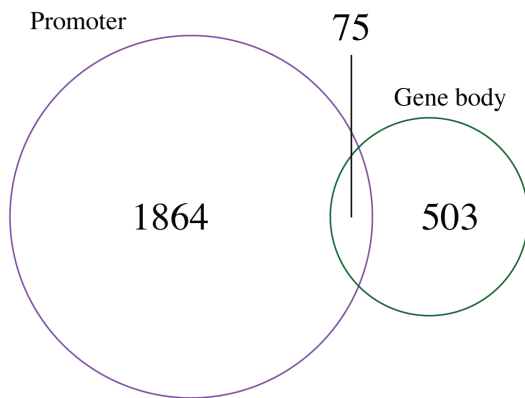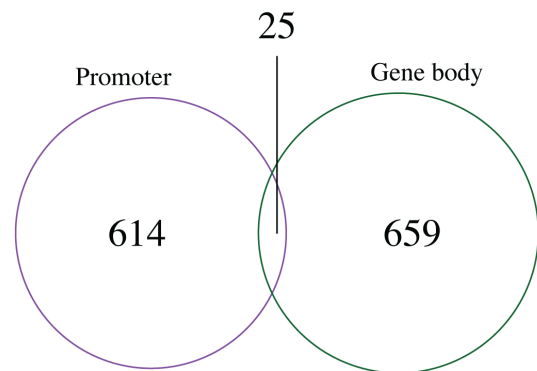**Supplementary Figure 1**. Initial differential DNA methylation analysis.

A) Volcano plot showing relationship between methylation difference and significance for differential promoter methylation analysis between all ASD samples and controls. A significance FDR cutoff of 0.1 is shown.

B) Volcano plot showing relationship between methylation difference and significance for differential gene body methylation analysis between all ASD samples and controls. A significance FDR cutoff of 0.1 is shown.

C) Relationship between sample loadings on the first principal component of differentially methylated promoters and gene bodies with technical and biological covariates. A Bonferroni corrected P-value of 0.05 is marked with a red line.

D) Relationship between sample loadings on the first principal component of differentially methylated promoters and gene bodies.
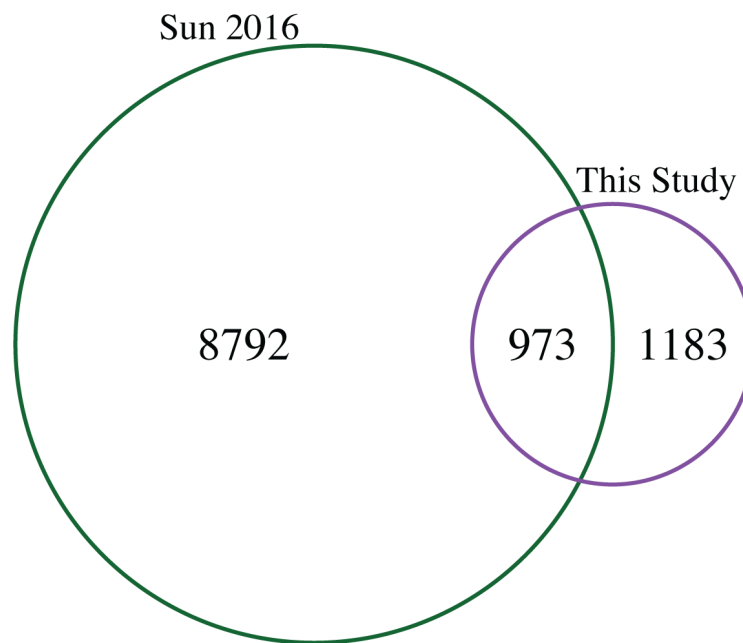
E) Overlap in genes between differentially methylated promoters and gene bodies that are hypermethylated in ASD.

F) Overlap in genes between differentially methylated promoters and gene bodies that are hypomethylated in ASD.

A.
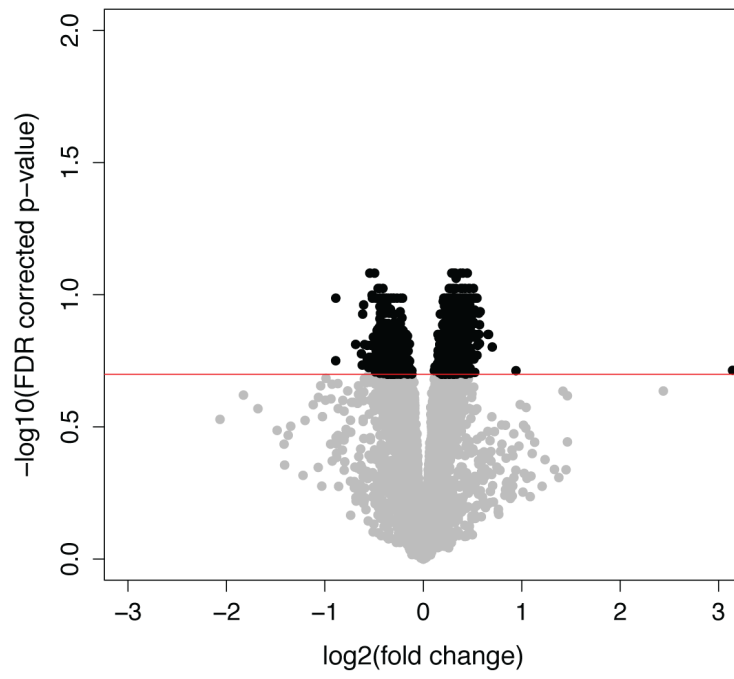
B.

**Supplementary Figure 2**. Initial differential histone acetylation analysis.

A) Volcano plot showing relationship between acetylation fold change and significance for differential histone acetylation analysis between all ASD samples and controls. A significance FDR cutoff of 0.2 is shown.

B) Overlap in differentially acetylated peaks between this analysis and Sun et al 2016.
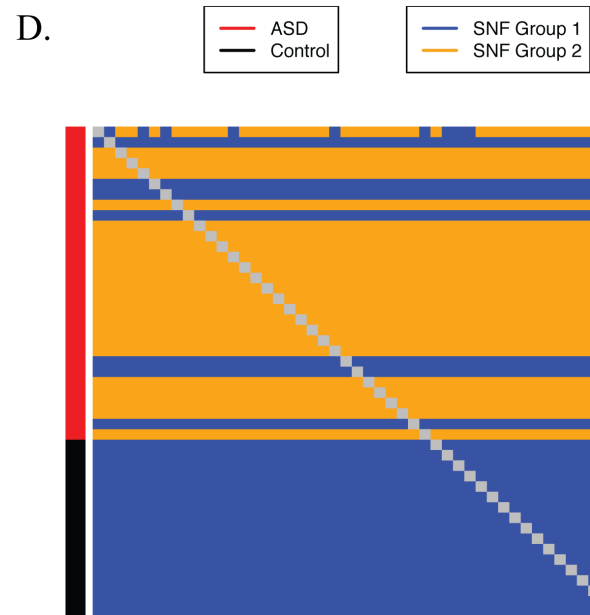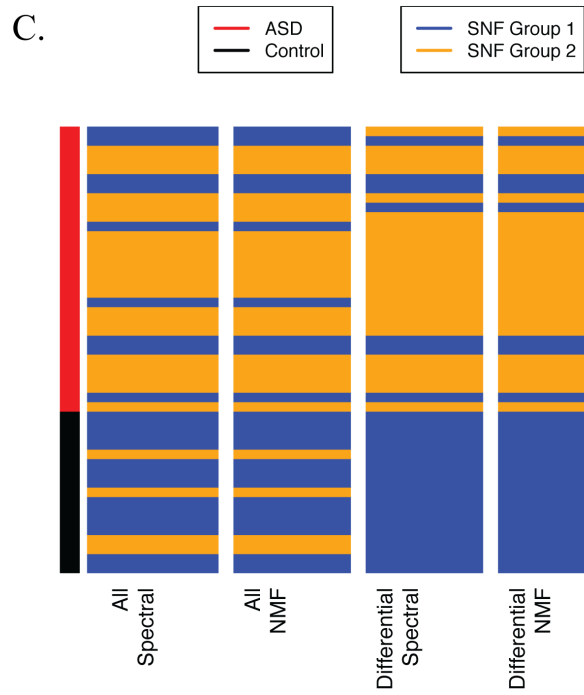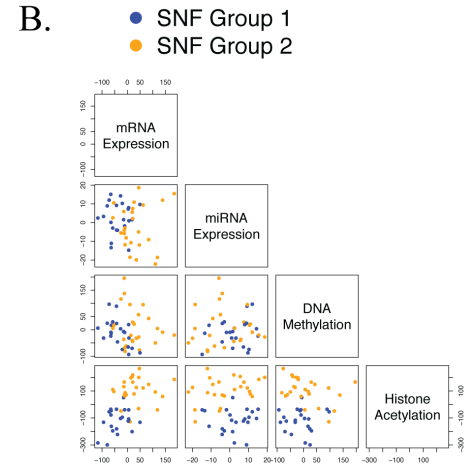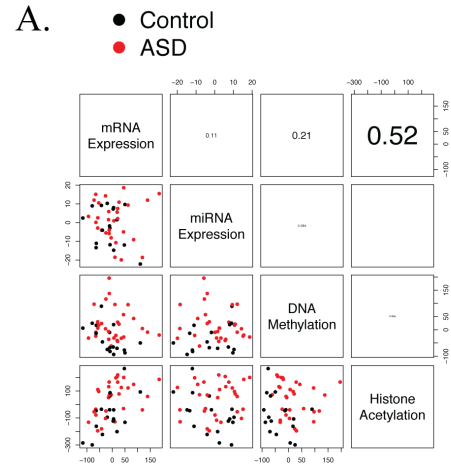
Supplementary Figure 3. SNF sample clustering.

A) Relationship between sample loadings on the first principal component of mRNA expression, miRNA expression, DNA methylation, and histone acetylation when utilizing all features from each dataset.

B) Classification of 2 sample clusters using SNF on all features from each of the 4 datasets.

C) Comparison of SNF cluster assignments when using all features or restricting to differential features as well as when using spectral clustering or nonnegative matrix factorization (NMF).

D) Comparison of SNF cluster assignments when exhaustively leaving each sample out and performing clustering on the remaining samples.

A.
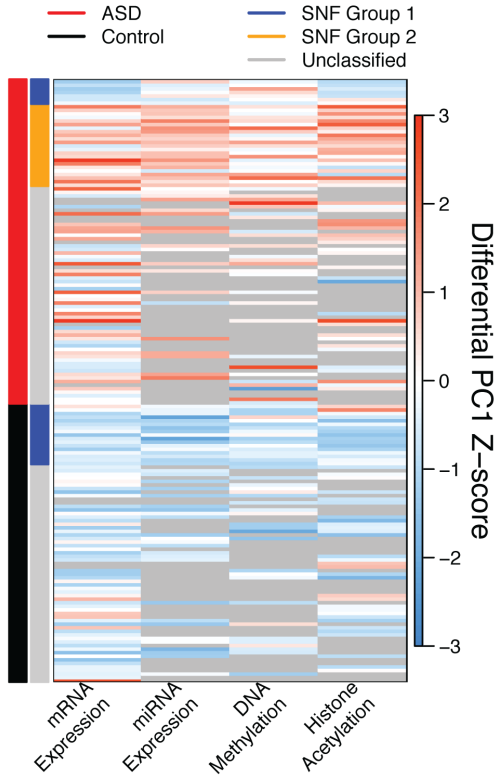
B.

C.

D.
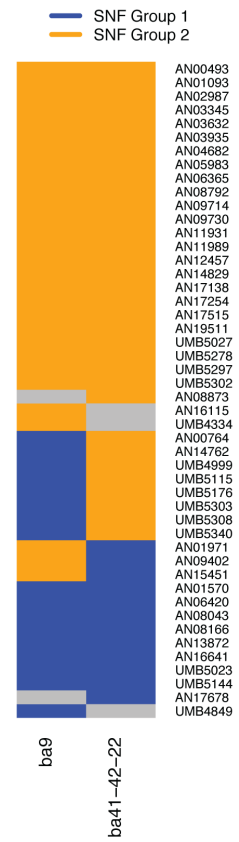
Supplementary Figure 4. Classification of samples into the 2 molecular subtypes.

A) Sample Z-score loadings for differential mRNA expression, miRNA expression, DNA methylation, and histone acetylation after running SNF, but before classification of samples that were missing in one or more datasets. The samples were sorted first by diagnosis, then by SNF subtype assignment. Samples not present in a particular dataset are colored in grey.

B) Exhaustive leave one out cross-validation accuracies for each logistic regression classification model. Two models (miRNA/Methylation/Acetylation and miRNA only) did not have any test samples and are colored black.

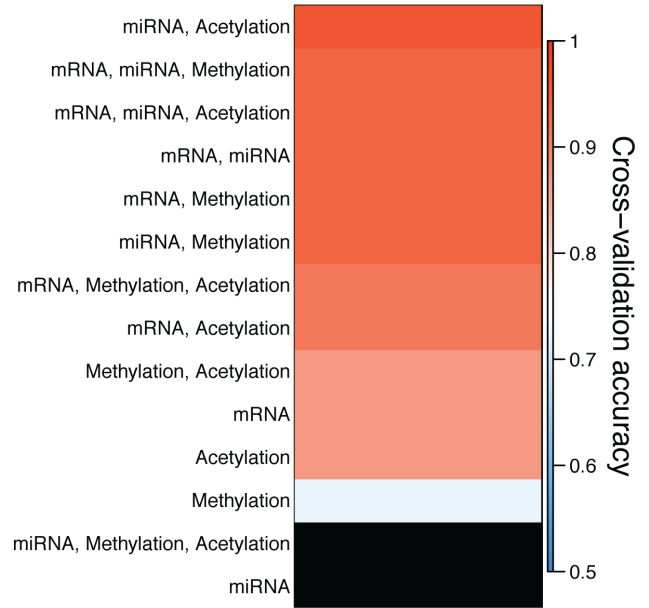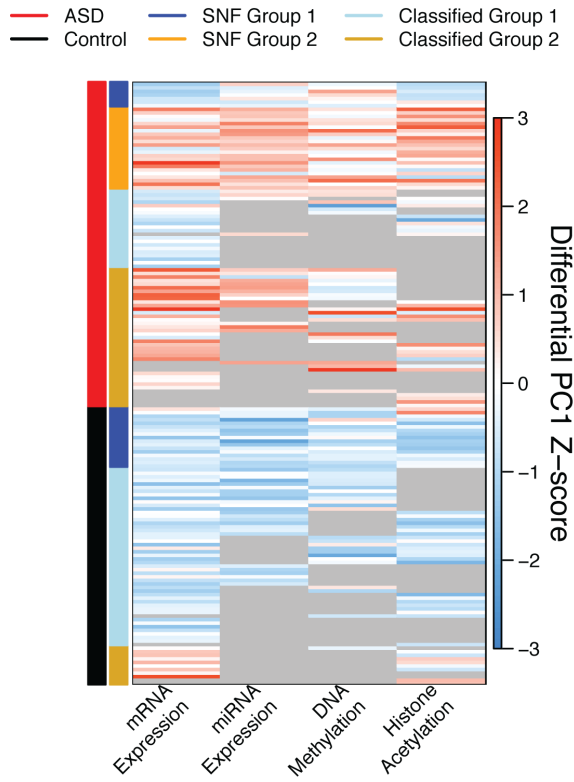**C**) Sample Z-score loadings for differential mRNA expression, miRNA expression, DNA methylation, and histone acetylation after classification of samples that were missing in one or more datasets. The samples were sorted first by diagnosis, then by SNF subtype assignment. Samples not present in a particular dataset are colored in grey.

D) Sample subtype assignments for 2 cortical regions from 48 ASD individuals. Individuals not present in any of the 4 datasets from a particular cortical region are colored in grey.
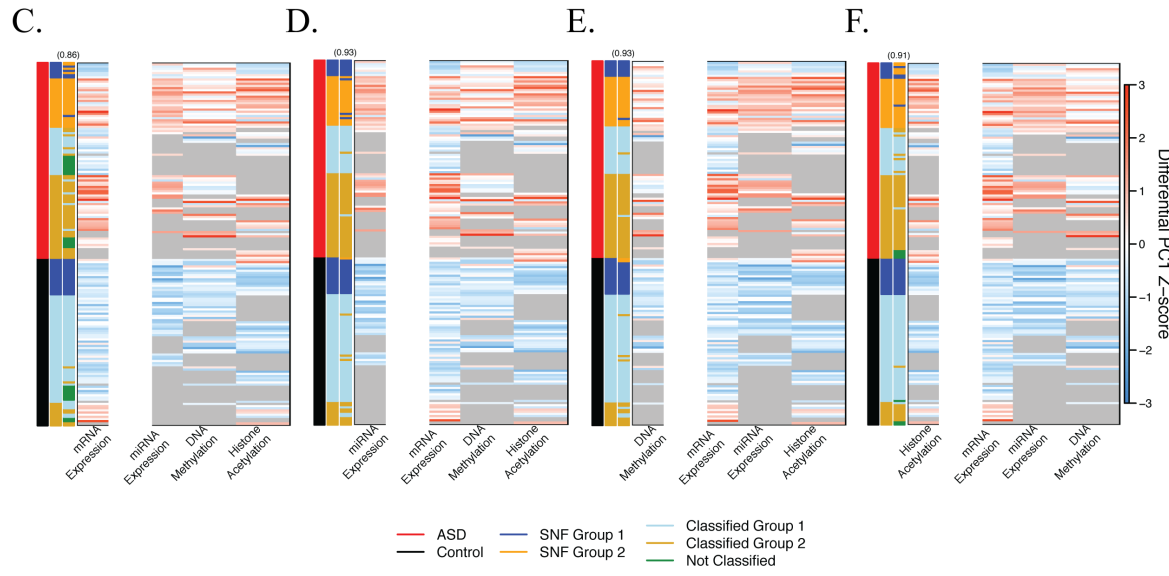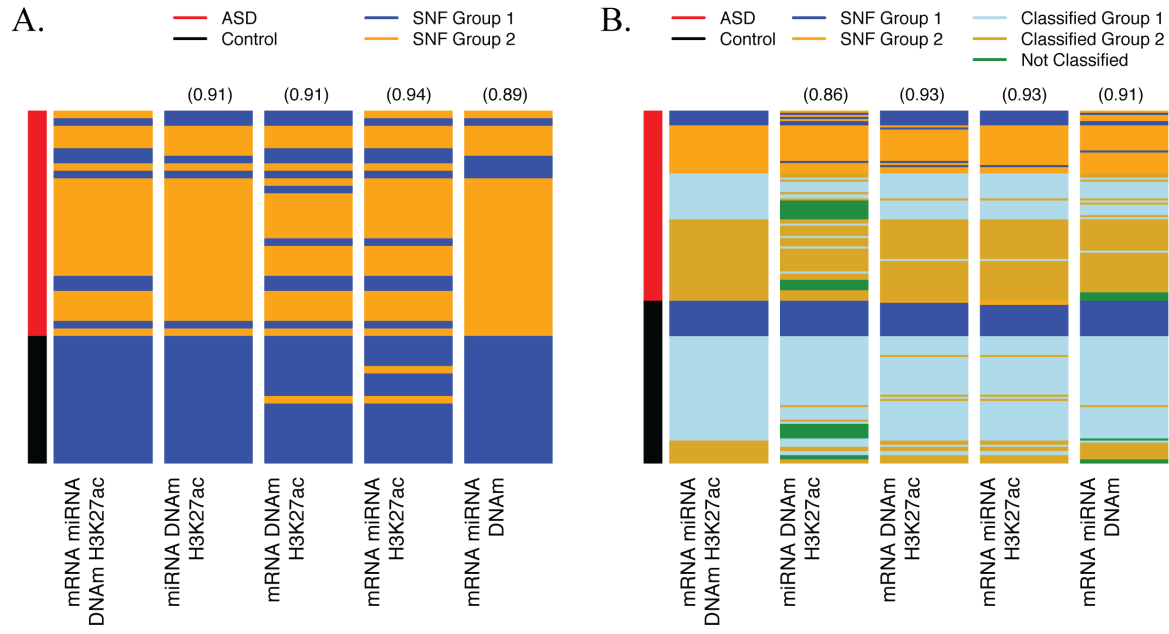
Supplementary Figure 5. Robustness of sample classifications.

A) Comparison of SNF clustering assignments for 47 samples using all of the four datasets to those when utilizing three out of four datasets. The samples are sorted by diagnosis. The concordances of sample assignments to those when using the complete dataset are shown in parentheses. The datasets used in SNF clustering are shown below each column.

B) Comparison of SNF clustering and classification assignments for 169 samples using all of the four datasets to those when utilizing three out of four datasets. The samples are sorted by diagnosis. The concordances of sample assignments to those when using the complete dataset are shown in parentheses. The datasets used in SNF clustering are shown below each column.

C) Sample Z-score loadings for differential mRNA expression, miRNA expression, DNA methylation, and histone acetylation after SNF clustering and classification of samples using miRNA expression, DNA methylation, and histone acetylation datasets while holding out the mRNA expression dataset (third y-axis color label). The samples were sorted first by diagnosis (first y-axis color label), then by subtype assignment when using all four datasets (second y-axis color label). The concordance between subtype assignments when using all four datasets as compared to those when using only three datasets is shown in parenthesis.

D) Same as **C**, except the held-out dataset is miRNA expression.

E) Same as **C**, except the held-out dataset is DNA methylation.

F) Same as **C**, except the held-out dataset is histone acetylation.

11

12
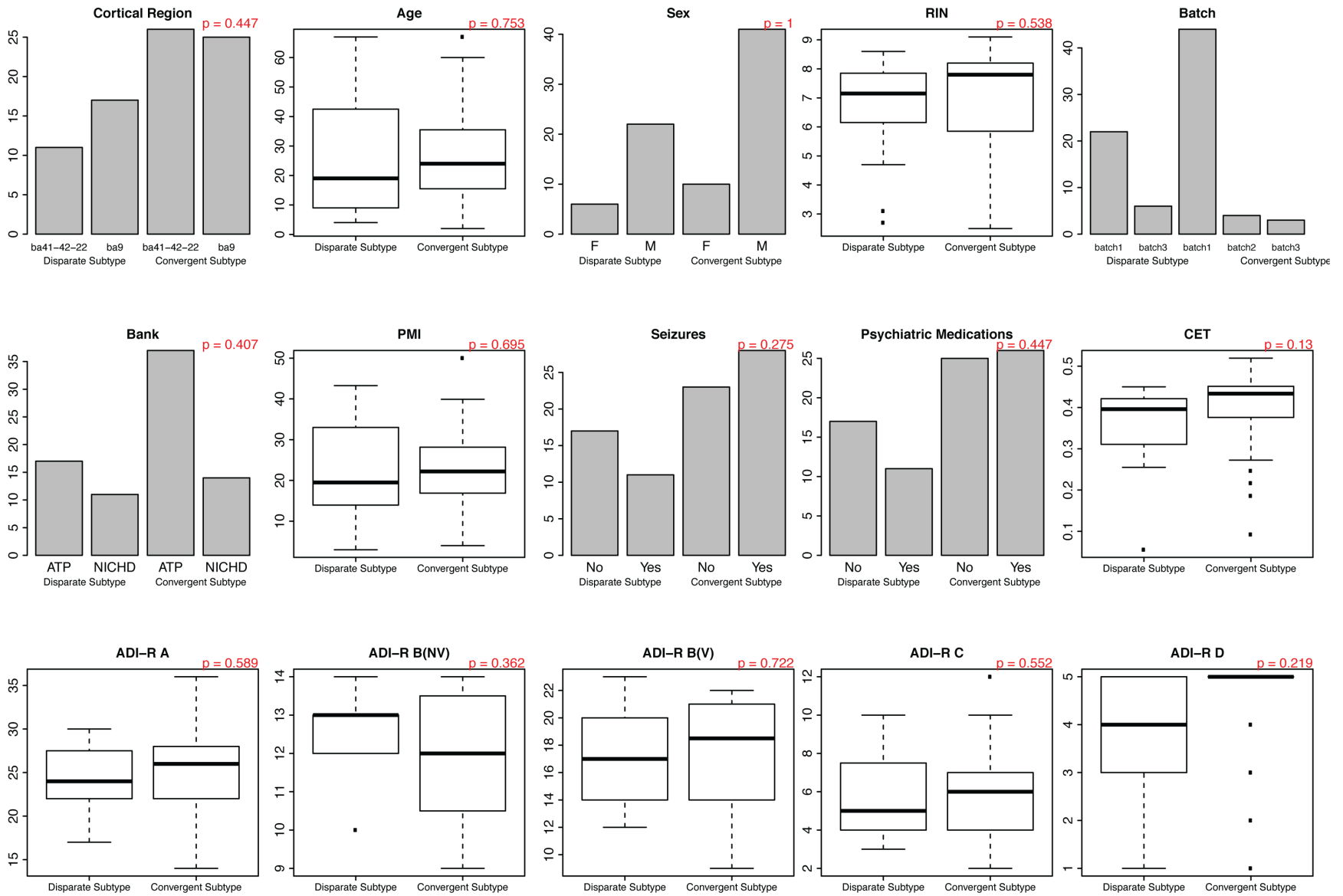
Supplementary Figure 6. Comparison of biological and technical covariates between the two

ASD sample subtypes. Significance of categorical covariates (Cortical region, Sex, Bank,

Seizures, and Psychiatric medications) was calculated using a two-tailed Chi-squared test.

Significance of numerical covariates (Age, RIN, PMI, ADI-R A, ADI-R B(NV), ADI-R B(V),

ADI-R C, ADI-R D, and CET) was calculated using a two-sided t-test. For boxplots of numerical

covariates, the center of the box is the median value, the bounds of the box are the 75th and 25th

percentile values, the whiskers extend out from the box to 1.5 times the interquartile range of the

box, and outlier values are presented as individual points.

A.

−log10(FDR corrected p−value)

log2(fold change)

B.

−log10(FDR corrected p−value)

log2(fold change)

C.

ASD up

- defense response
- response to cytokine
- regulation of immune system process
- immune effector process
- cytokine production

−log10(FDR corrected p−value)

D.

ASD down

- trans−synaptic signaling
- ion transmembrane transport
- inorganic molecular entity transmembrane transporter activity
- metal ion transport
- regulation of ion transport

−log10(FDR corrected p−value)

E.

Module eigengene associations with diagnosis and covariates

	ASD−CTL	Age	Sex	Region	Batch2	Batch3	RIN	Bank	SeqPC1	SeqPC2	SeqPC3	SeqPC4	SeqPC5
mRNA.M1	0.0036	1.9e−11											
mRNA.M2													
mRNA.M3	4e−04												
mRNA.M4	1.6e−05	0.04											
mRNA.M5		0.013											
mRNA.M6		0.039											
mRNA.M7	0.0047	2.3e−05											
mRNA.M8	2.4e−07			0.01									
mRNA.M9	8.8e−06												
mRNA.M10	1.9e−05	0.00025											
mRNA.M12													
mRNA.M13													
mRNA.M14													
mRNA.M15	0.00027												
mRNA.M16	2e−05												
mRNA.M17	0.0012												
mRNA.M18		0.027											
mRNA.M19	0.00015									0.017			
mRNA.M20	0.018	7.3e−11											
mRNA.M21	0.0019												
mRNA.M22		3e−08											
mRNA.M23	0.00034									0.026			
mRNA.M24													
mRNA.M25		0.011											

log10(signed Bonferroni corrected p−value)

F.

R = 0.91
P = 0
Slope = 1.1

log2(fold change) − Current study

log2(fold change) − Parikshak 2016

G.

Cell type enrichment

	Oligo−dendrocytes	Microglia	Astrocyte	Endothelial	Neuron
mRNA.M1					
mRNA.M2	5.6 (3.2e−88)				
mRNA.M3					2.4 (5.7e−15)
mRNA.M4					3.5 (9.1e−10)
mRNA.M5	3.2 (0.00063)				
mRNA.M6					
mRNA.M7					
mRNA.M8					
mRNA.M9		9.4 (6e−92)	1.9 (7e−04)		
mRNA.M10					2.9 (1.6e−06)
mRNA.M12					
mRNA.M13					
mRNA.M14			3.5 (8.1e−06)		
mRNA.M15		3.9 (1.5e−06)		2.4 (0.019)	
mRNA.M16					4.2 (1e−15)
mRNA.M17					4.6 (2.3e−14)
mRNA.M18		11 (7e−66)		5.1 (1.1e−23)	
mRNA.M19		8.4 (2.1e−44)		4.2 (5.7e−15)	
mRNA.M20					
mRNA.M21					
mRNA.M22					
mRNA.M23			2.2 (0.0022)		
mRNA.M24					
mRNA.M25					3.6 (6.9e−28)

Odds ratio

Supplementary Figure 7. mRNA expression differences in ASD.

A) Volcano plot showing relationship between expression fold change and significance for differential expression analysis between ASD Convergent subtype samples and controls. A significance FDR cutoff of 0.05 is shown.

B) Volcano plot showing relationship between expression fold change and significance for differential expression analysis between ASD Disparate subtype samples and controls. A significance FDR cutoff of 0.05 is shown.

C) Top gene ontology enrichments for upregulated genes in ASD. Ontology enrichments were calculated by g:Profiler with FDR corrected p-values.

D) Top gene ontology enrichments for downregulated genes in ASD. Ontology enrichments were calculated by g:Profiler with FDR corrected p-values.

E) Module-trait associations as computed by a linear mixed model with all factors on the x-axis used as covariates. P-values were adjusted for multiple testing using Bonferroni correction. Only p-values with a Bonferroni-corrected value < 0.05 are shown.

F) Comparison of expression fold changes between ASD Convergent subtype vs control analysis in this study and idiopathic ASD vs control analysis of Parikshak et al[9]. P-value is from a linear model.

G) Cell type enrichments for all of the mRNA co-expression modules. Enrichments were calculated using a logistic regression model and p-values, which are shown in parentheses, were adjusted for multiple testing using FDR correction. Only those enrichments with odds ratio > 1 and FDR corrected p-value < 0.05 are shown.

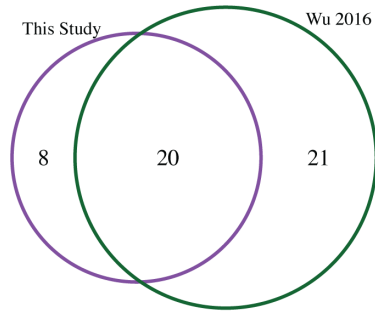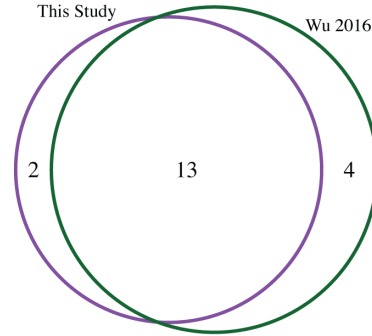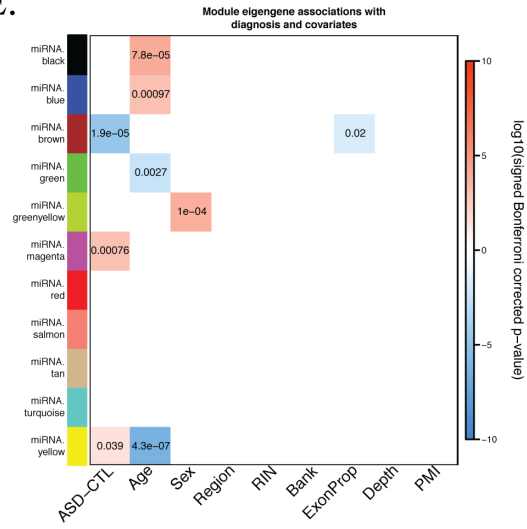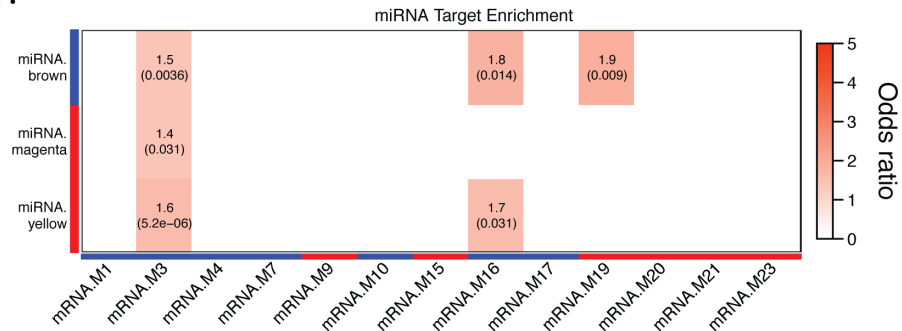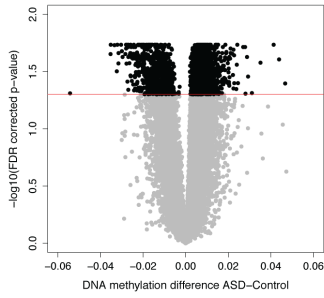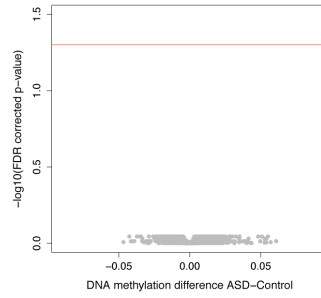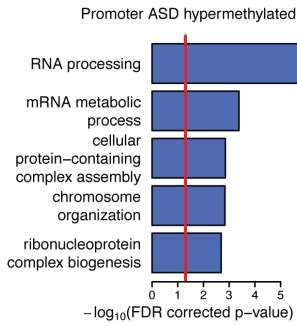Source data are provided as a Source Data file.
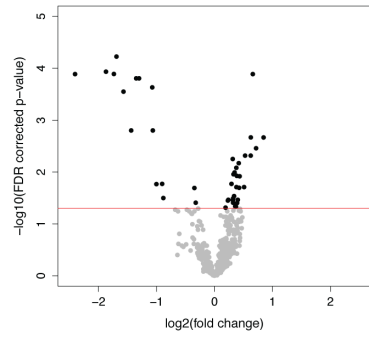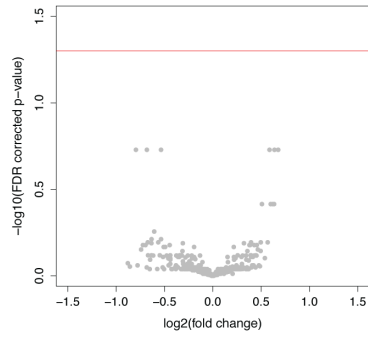
Supplementary Figure 8. miRNA expression differences in ASD.

A) Volcano plot showing relationship between miRNA expression fold change and significance for differential miRNA expression analysis between ASD Convergent subtype samples and controls. A significance FDR cutoff of 0.05 is shown.

B) Volcano plot showing relationship between miRNA expression fold change and significance for differential miRNA expression analysis between ASD Disparate subtype samples and controls. A significance FDR cutoff of 0.05 is shown.

C) Overlap in miRNAs that are significantly upregulated in ASD found in this study compared to Wu et al 2016.

D) Overlap in miRNAs that are significantly downregulated in ASD found in this study compared to Wu et al 2016.

E) Module-trait associations as computed by a linear mixed model with all factors on the x-axis used as covariates. P-values were adjusted for multiple testing using Bonferroni correction. Only p-values with a Bonferroni-corrected value < 0.05 are shown.

F) Enrichment of differentially expressed mRNA co-expression modules with predicted targets for constituents of each differentially expressed miRNA co-expression module. Module relationships to diagnosis are marked along the x and y axes (red: increased expression in ASD; blue: decreased expression in ASD). Enrichments were calculated using a logistic regression model and p-values, which are shown in parentheses, were adjusted for multiple testing using FDR correction. Only those enrichments with odds ratio > 1 and FDR corrected p-value < 0.05 are shown.
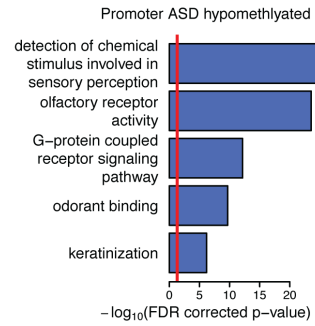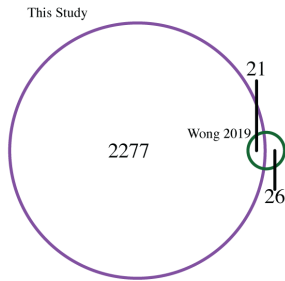
Source data are provided as a Source Data file.

A. Promoter Co-methylation network dendrogram

B. C.

D. Promoter ASD hypermethylated

E. Promoter ASD hypomethlyated

F. Module eigengene associations with diagnosis and covariates

G. Cell type enrichment

H. This Study

I. This Study
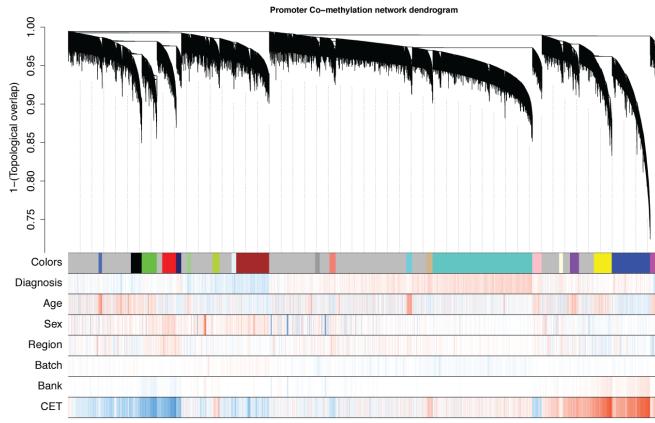
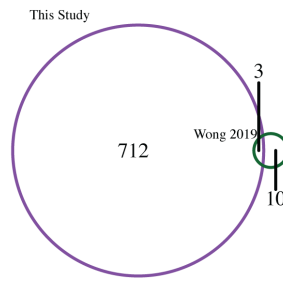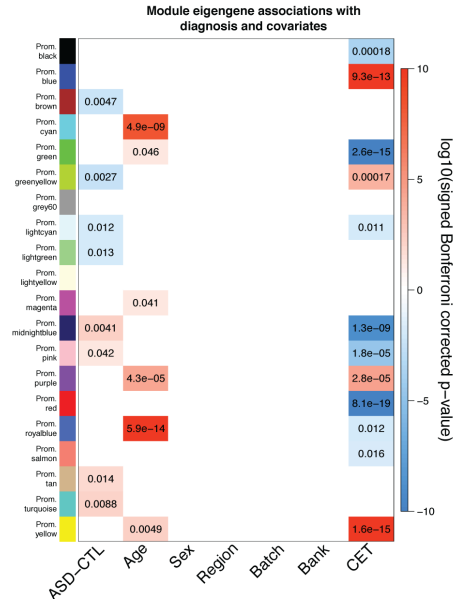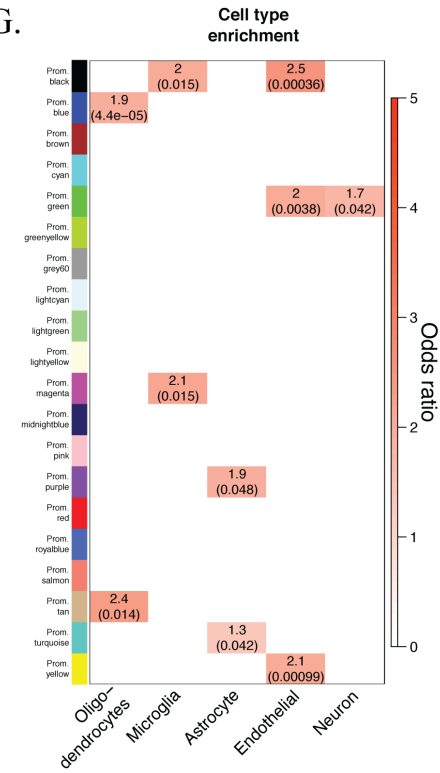**Supplementary Figure 9**. Promoter DNA methylation differences in ASD.

A) Average linkage hierarchical clustering using the topological overlap metric for promoter co-methylation dissimilarity. Modules were identified from this dendrogram, which was constructed from a consensus of 100 bootstrapped datasets. Correlations for each gene to covariates are delineated below the dendrogram (blue, negative; red, positive). Modules are labelled with colors. The grey module represents genes that are not co-methylated and was not evaluated in further comparisons.

B) Volcano plot showing relationship between methylation change and significance for differential promoter methylation analysis between ASD Convergent subtype samples and controls. A significance FDR cutoff of 0.05 is shown.

C) Volcano plot showing relationship between methylation change and significance for differential promoter methylation analysis between ASD Disparate subtype samples and controls. A significance FDR cutoff of 0.05 is shown.

D) Top gene ontology enrichments for genes with hypermethylated promoters in ASD. Ontology enrichments were calculated by g:Profiler with FDR corrected p-values.

E) Top gene ontology enrichments for genes with hypomethylated promoters in ASD. Ontology enrichments were calculated by g:Profiler with FDR corrected p-values.

F) Module-trait associations as computed by a linear mixed model with all factors on the x-axis used as covariates. P-values were adjusted for multiple testing using Bonferroni correction. Only p-values with a Bonferroni-corrected value < 0.05 are shown.

G) Cell type enrichments for all of the promoter co-methylation modules. Enrichments were calculated using a logistic regression model and p-values, which are shown in parentheses, were
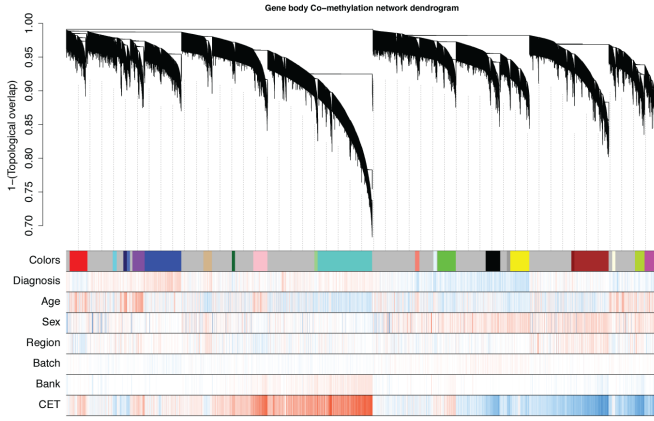
adjusted for multiple testing using FDR correction. Only those enrichments with odds ratio > 1 and FDR corrected p-value < 0.05 are shown.

H) Overlap in ASD hypermethylated gene promoters identified in this study with gene promoters containing ASD hypermethylated probes identified in Wong et al 2019.
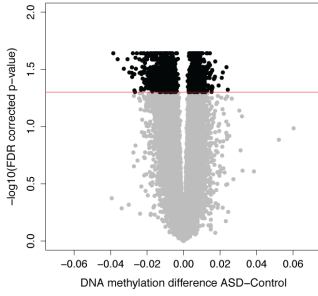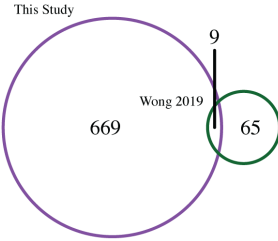
I) Overlap in ASD hypomethylated gene promoters identified in this study with gene promoters containing ASD hypomethylated probes identified in Wong et al 2019.

Source data are provided as a Source Data file.

A. Gene body Co-methylation network dendrogram

B.

C.

D. Gene body ASD hypermethylated

E. Gene body ASD hypomethylated

F. Module eigengene associations with diagnosis and covariates

G. Cell type enrichment

H. This Study

I. This Study
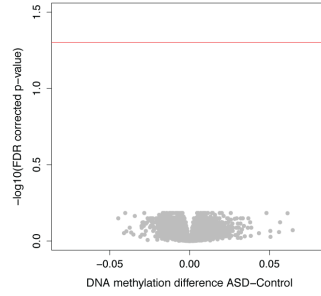
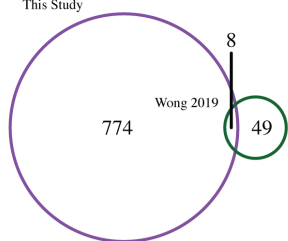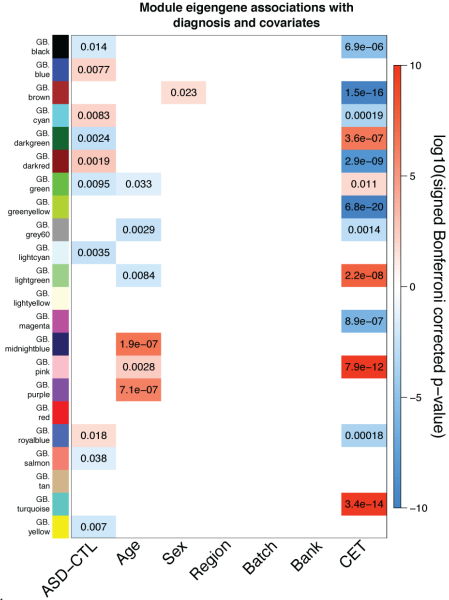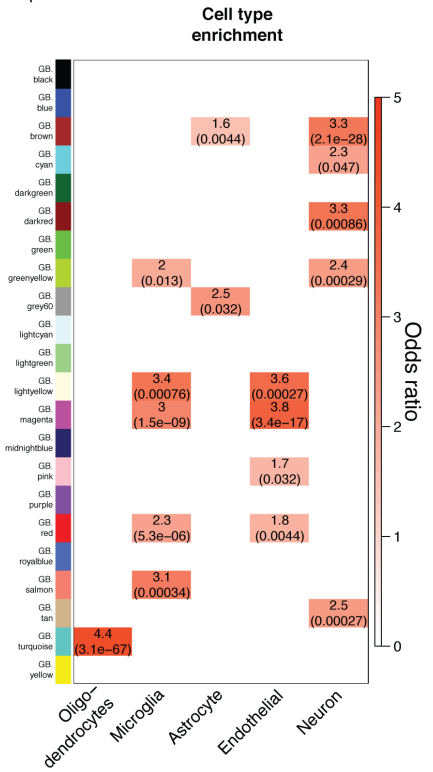**Supplementary Figure 10**. Gene body DNA methylation differences in ASD.

A) Average linkage hierarchical clustering using the topological overlap metric for gene body co-methylation dissimilarity. Modules were identified from this dendrogram, which was constructed from a consensus of 100 bootstrapped datasets. Correlations for each gene to covariates are delineated below the dendrogram (blue, negative; red, positive). Modules are labelled with colors. The grey module represents genes that are not co-methylated and was not evaluated in further comparisons.

B) Volcano plot showing relationship between methylation change and significance for differential gene body methylation analysis between ASD Convergent subtype samples and controls. A significance FDR cutoff of 0.05 is shown.

C) Volcano plot showing relationship between methylation change and significance for differential gene body methylation analysis between ASD Disparate subtype samples and controls. A significance FDR cutoff of 0.05 is shown.

D) Top gene ontology enrichments for hypermethylated gene bodies in ASD. Ontology enrichments were calculated by g:Profiler with FDR corrected p-values.

E) Top gene ontology enrichments for hypomethylated gene bodies in ASD. Ontology enrichments were calculated by g:Profiler with FDR corrected p-values.

F) Module-trait associations as computed by a linear mixed model with all factors on the x-axis used as covariates. P-values were adjusted for multiple testing using Bonferroni correction. Only p-values with a Bonferroni-corrected value < 0.05 are shown.

G) Cell type enrichments for all of the gene body co-methylation modules. Enrichments were calculated using a logistic regression model and p-values, which are shown in parentheses, were
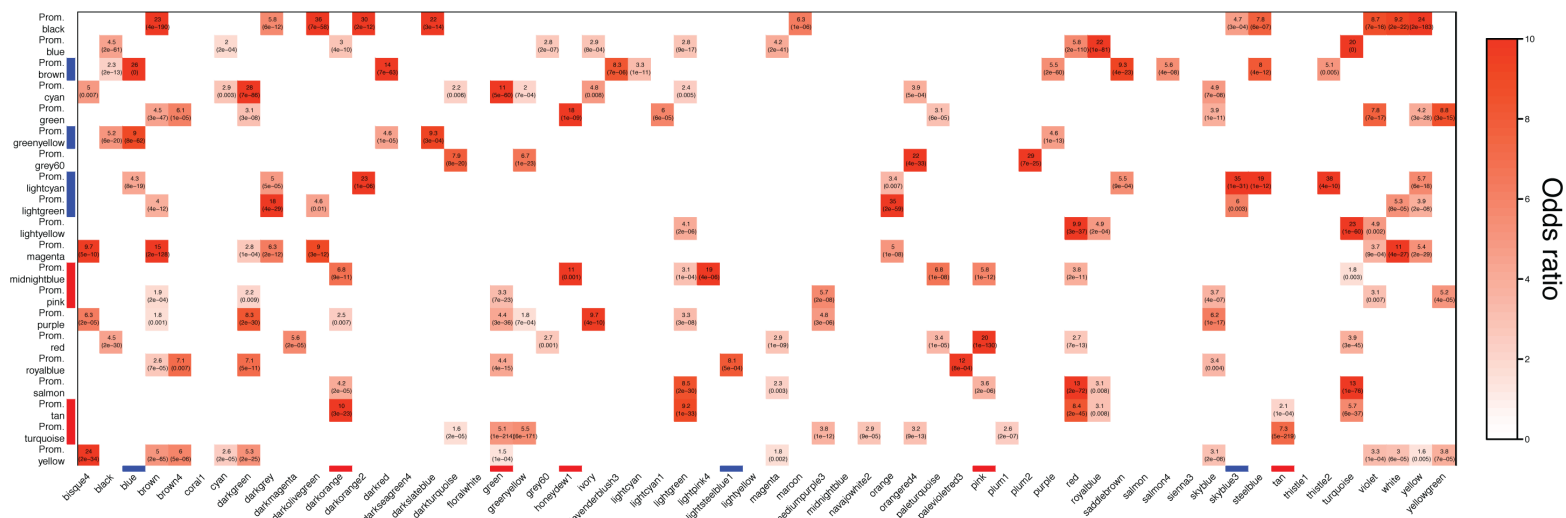
adjusted for multiple testing using FDR correction. Only those enrichments with odds ratio > 1 and FDR corrected p-value < 0.05 are shown.

H) Overlap in ASD hypermethylated gene bodies identified in this study with gene bodies containing ASD hypermethylated probes identified in Wong et al 2019.
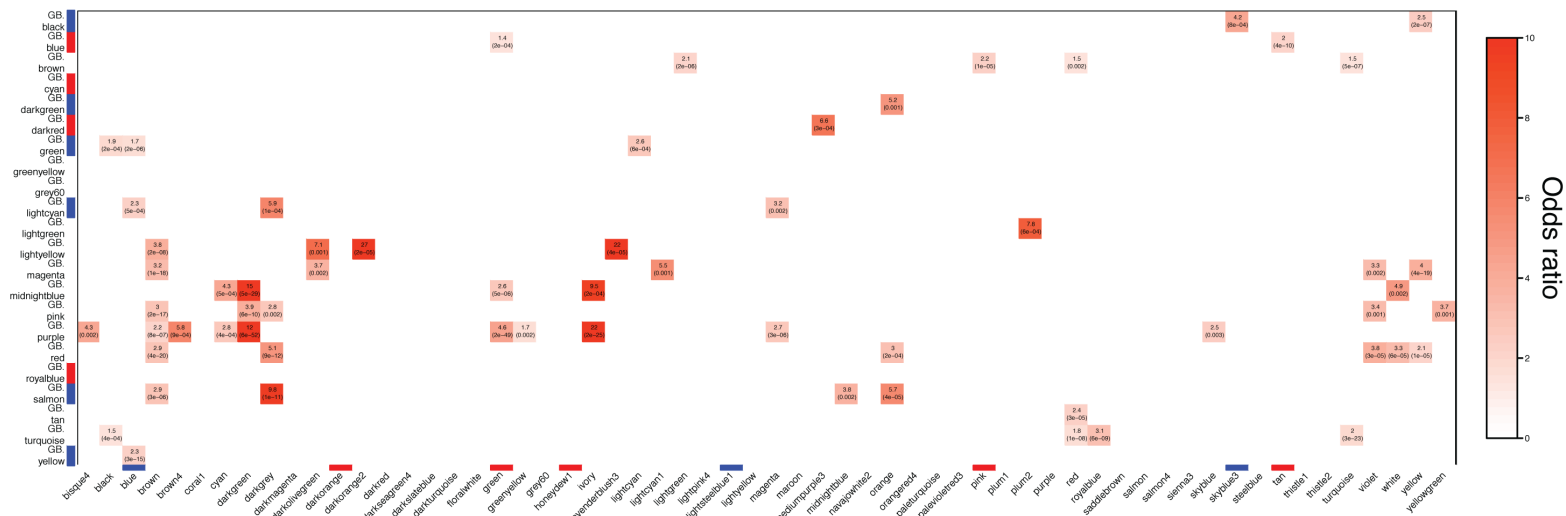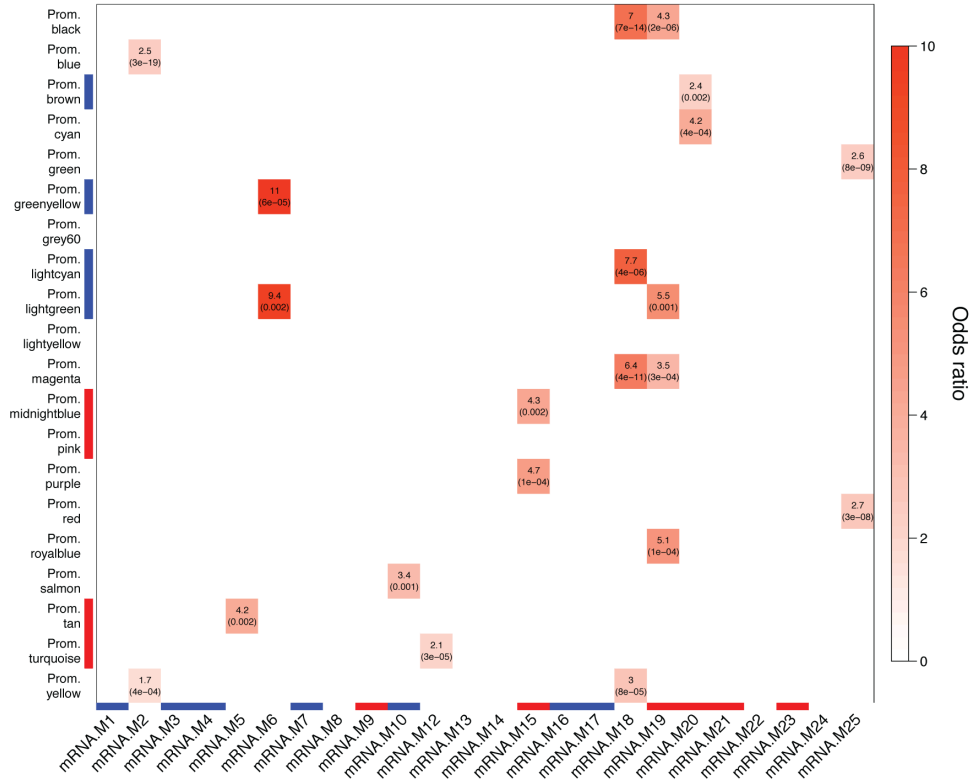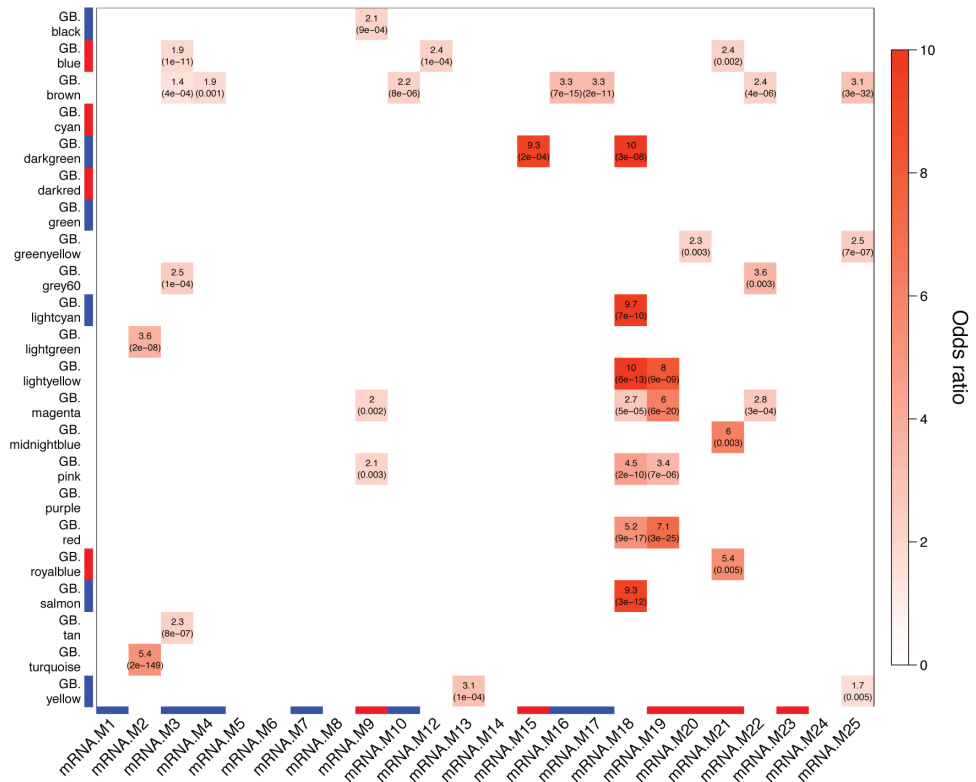
I) Overlap in ASD hypomethylated gene bodies identified in this study with gene bodies containing ASD hypomethylated probes identified in Wong et al 2019.

Source data are provided as a Source Data file.

C.

D.

Supplementary Figure 11. Comparison of gene level co-methylation networks to previously published probe level co-methylation network and co-expression network.
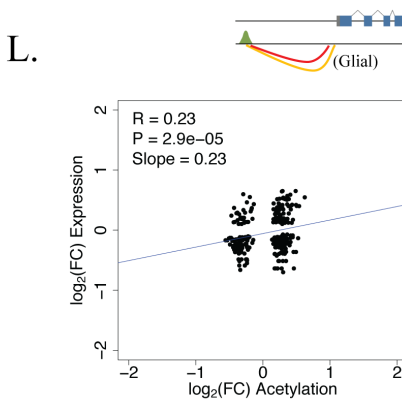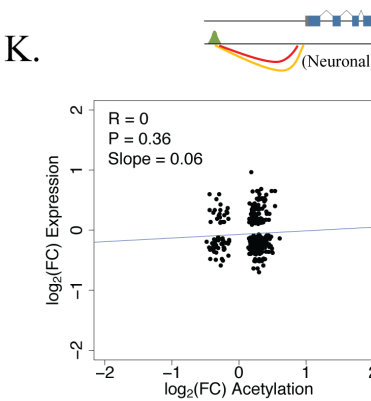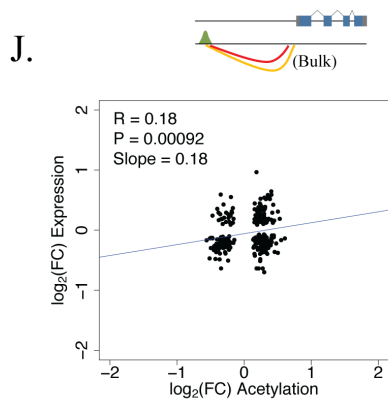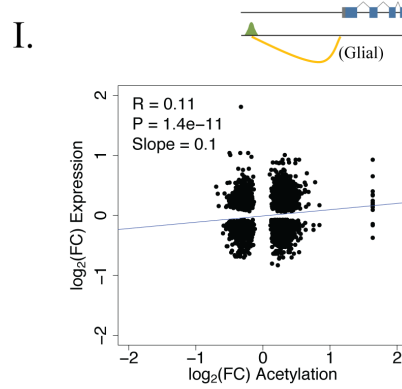
A) Comparison of promoter co-methylation modules to probe level co-methylation modules identified in Wong et al 2019. Modules with a significant relationship to diagnosis are marked along the x and y axes (red: hypermethylated in ASD; blue: hypomethylated in ASD). Enrichments were calculated using a logistic regression model and p-values were adjusted for multiple testing using FDR correctio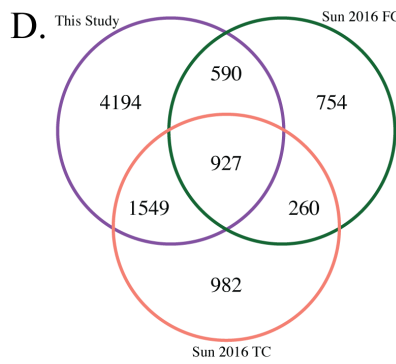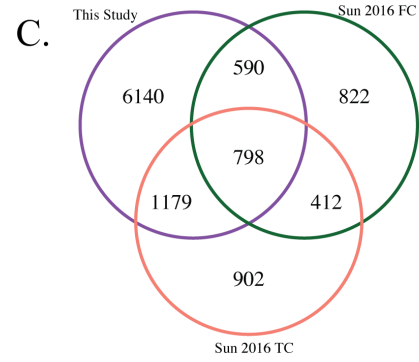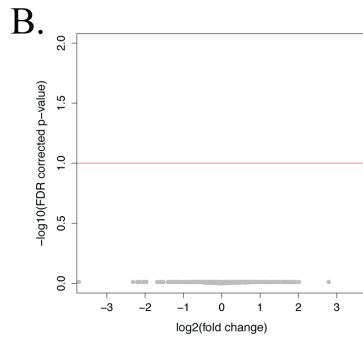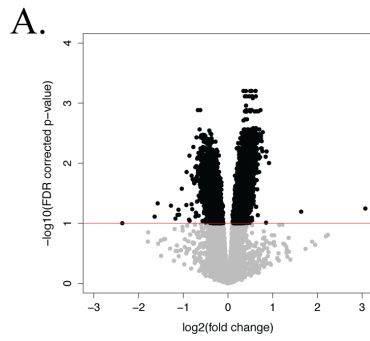n. Uncorrected p-values are shown in parentheses. Only those enrichments with odds ratio > 1 and FDR corrected p-value < 0.05 are shown.

B) Same as **A** except for gene body co-methylation modules.

C) Comparison of promoter co-methylation modules to mRNA co-expression modules. Modules with a significant relationship to diagnosis are marked along the x and y axes (red: hypermethylated or increased expression in ASD; blue: hypomethylated or decreased expression in ASD). Enrichments were calculated using a logistic regression model and p-values were adjusted for multiple testing using FDR correction. Uncorrected p-values are shown in parentheses. Only those enrichments with odds ratio > 1 and FDR corrected p-value < 0.05 are shown.

D) Same as **C** except for gene body co-methylation modules.

A.

B.

C. This Study — Sun 2016 FC

6140 590 822
798
1179 412
902
Sun 2016 TC

D. This Study — Sun 2016 FC

4194 590 754
927
1549 260
982
Sun 2016 TC

E.
R = 0.018
P = 2e−08
Slope = 0.0

F.
R = 0.084
P = 2.3e−09
Slope = 0.085

G. (Bulk)
R = 0.075
P = 1.4e−05
Slope = 0.082

H. (Neuronal)
R = 0.061
P = 7.6e−05
Slope = 0.067

I. (Glial)
R = 0.11
P = 1.4e−11
Slope = 0.1

J. (Bulk)
R = 0.18
P = 0.00092
Slope = 0.18

K. (Neuronal)
R = 0
P = 0.36
Slope = 0.06

L. (Glial)
R = 0.23
P = 2.9e−05
Slope = 0.23

27
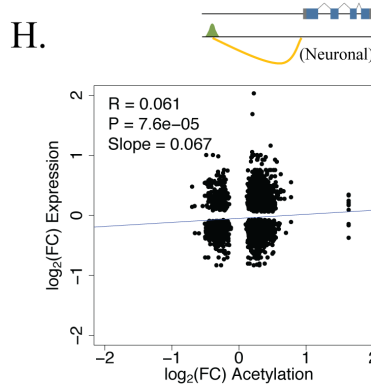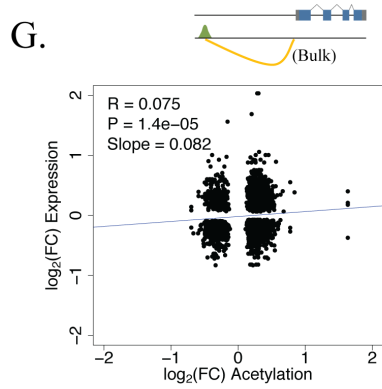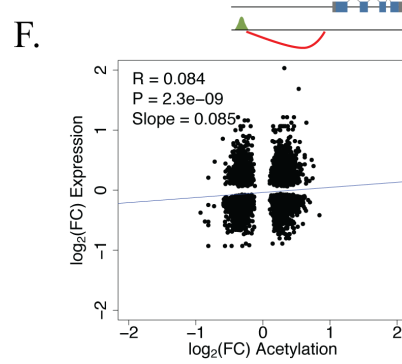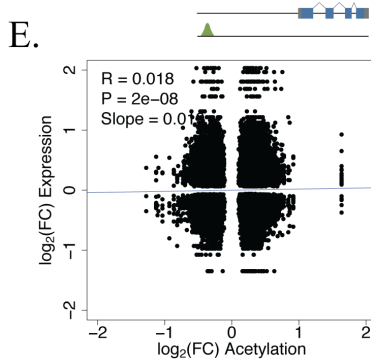
Supplementary Figure 12. Histone acetylation differences in ASD.

A) Volcano plot showing relationship between acetylation fold change and significance for differential acetylation analysis between ASD Convergent subtype samples and controls. A significance FDR cutoff of 0.1 is shown.

B) Volcano plot showing relationship between acetylation fold change and significance for differential acetylation analysis between ASD Disparate subtype samples and controls. A significance FDR cutoff of 0.1 is shown.

C) Overlap in ASD hyperacetylated regions between this study and Sun et al 2016.

D) Overlap in ASD hypoacetylated regions between this study and Sun et al 2016.

E) Correlation between expression and acetylation changes for differentially acetylated regions within 1 MB of the TSS for differentially expressed genes. P-value is from a linear model used to correlate differential expression with differential acetylation.

F) Correlation between expression and acetylation changes for differentially acetylated regions linked to differentially expressed genes with eQTL evidence. P-value is from a linear model used to correlate differential expression with differential acetylation.

G) Correlation between expression and acetylation changes for differentially acetylated regions linked to differentially expressed genes with bulk tissue Hi-C evidence. P-value is from a linear model used to correlate differential expression with differential acetylation.

H) Correlation between expression and acetylation changes for differentially acetylated regions linked to differentially expressed genes with neuronal Hi-C evidence. P-value is from a linear model used to correlate differential expression with differential acetylation.

I) Correlation between expression and acetylation changes for differentially acetylated regions linked to differentially expressed genes with glial Hi-C evidence. P-value is from a linear model used to correlate differential expression with differential acetylation.

J) Correlation between expression and acetylation changes for differentially acetylated regions linked to differentially expressed genes with both eQTL and bulk tissue Hi-C evidence. P-value is from a linear model used to correlate differential expression with differential acetylation.

K) Correlation between expression and acetylation changes for differentially acetylated regions linked to differentially expressed genes with both eQTL and neuronal Hi-C evidence. P-value is from a linear model used to correlate differential expression with differential acetylation.

L) Correlation between expression and acetylation changes for differentially acetylated regions linked to differentially expressed genes with both eQTL and glial Hi-C evidence. P-value is from a linear model used to correlate differential expression with differential acetylation.
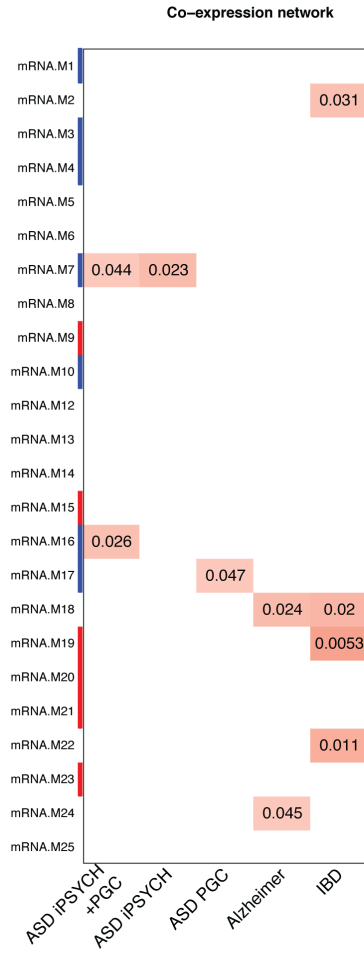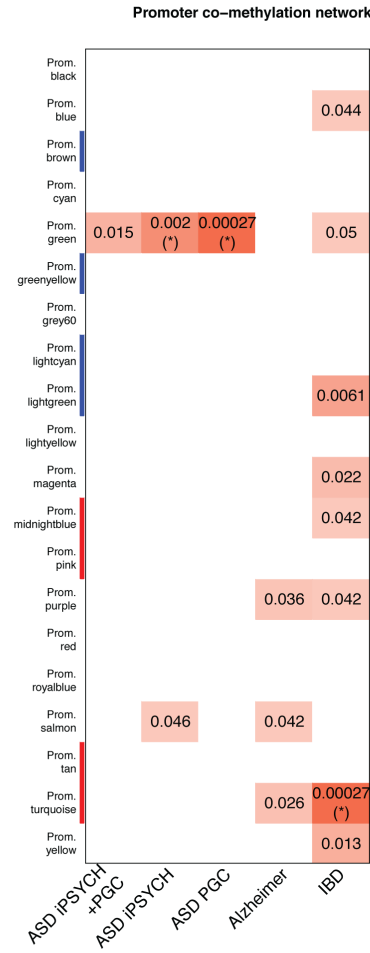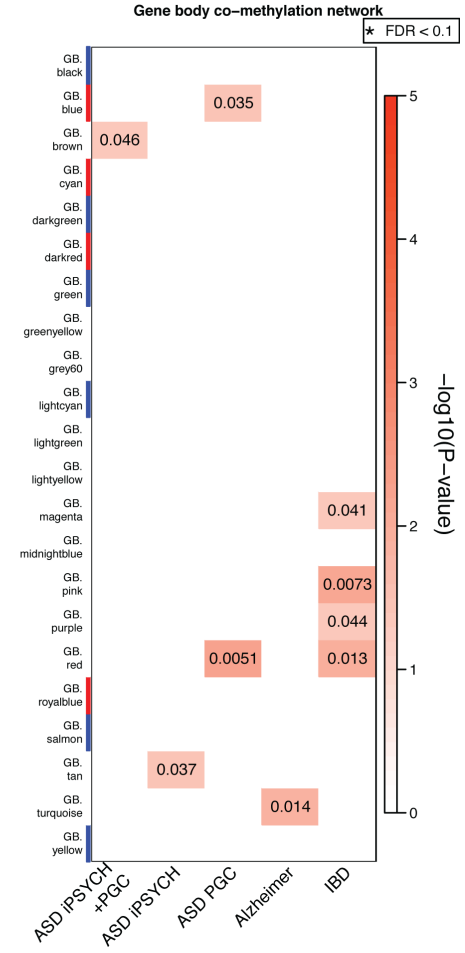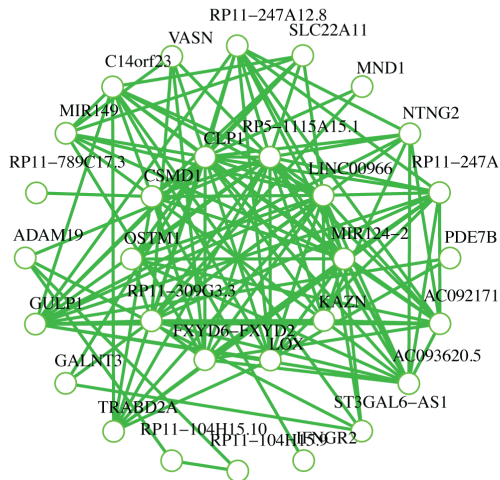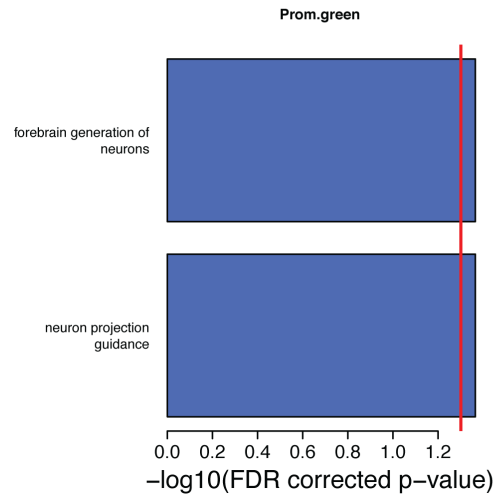
Supplementary Figure 13. ASD genetic risk enrichments in networks.

A) Partitioned heritability enrichments for ASD, Alzheimer's, and IBD GWAS in mRNA co-expression modules. Modules upregulated and downregulated in ASD are marked in red and blue, respectively. Uncorrected p-values < 0.05 are shown. P-values were also adjusted for multiple testing by FDR correction and those enrichments with adjusted p-value < 0.1 are marked with asterisks.

B) Partitioned heritability enrichments for ASD, Alzheimer's, and IBD GWAS in promoter co-methylation modules. Modules hypermethylated and hypomethylated in ASD are marked in red and blue, respectively. Uncorrected p-values < 0.05 are shown. P-values were also adjusted for multiple testing by FDR correction and those enrichments with adjusted p-value < 0.1 are marked with asterisks.

C) Partitioned heritability enrichments for ASD, Alzheimer's, and IBD GWAS in gene body co-methylation modules. Modules hypermethylated and hypomethylated in ASD are marked in red and blue, respectively. Uncorrected p-values < 0.05 are shown. P-values were also adjusted for multiple testing by FDR correction and those enrichments with adjusted p-value < 0.1 are marked with asterisks.
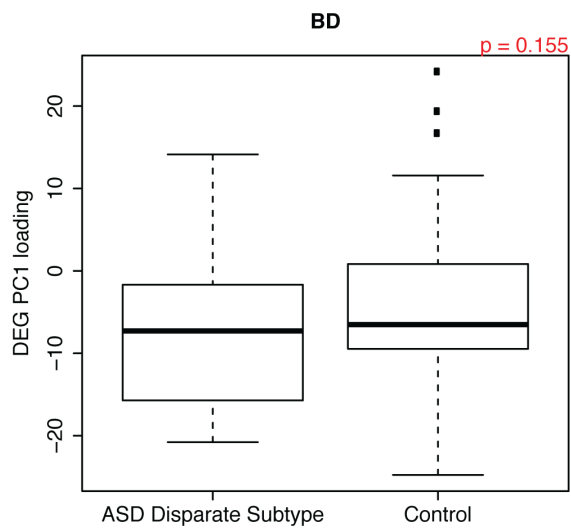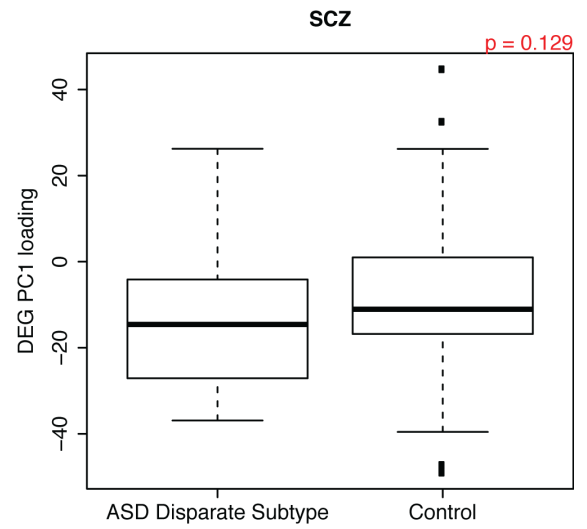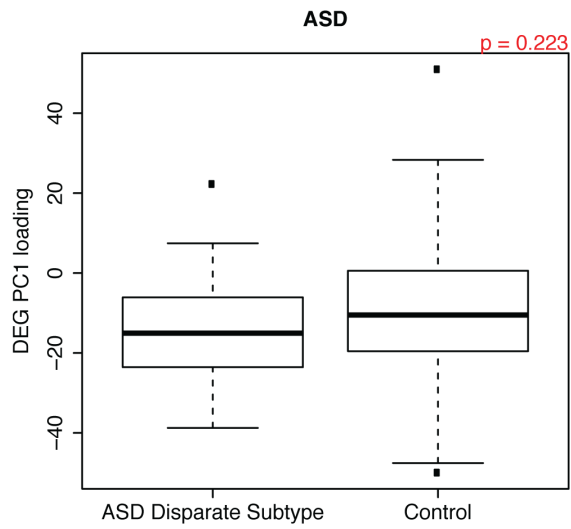
**D**) Top 30 hub genes and 300 connections for promoter co-methylation module Prom.green.

E) Top gene ontology enrichments for promoter co-methylation module Prom.green. Ontology enrichments were calculated by g:Profiler with FDR corrected p-values.

Source data are provided as a Source Data file.
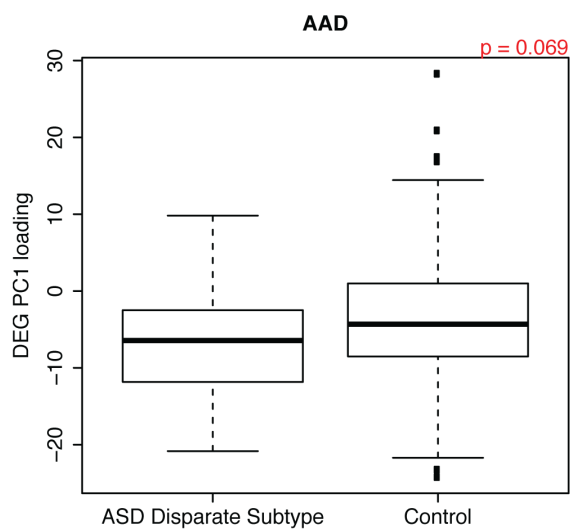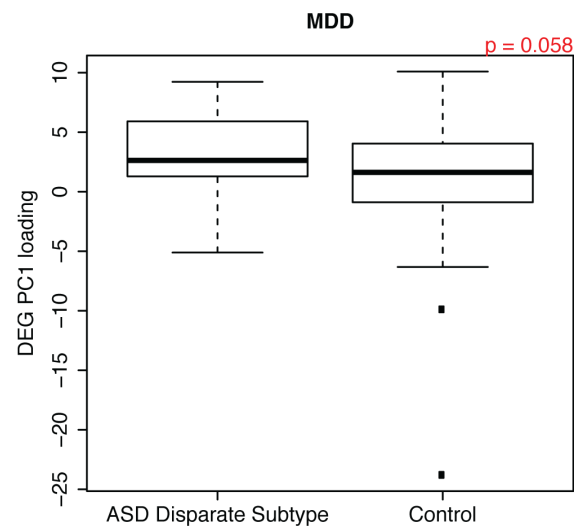
Supplementary Figure 14. Assessment of ASD Disparate subtype samples with respect to other neuropsychiatric disorders. Loadings on the first principal component when restricting ASD Disparate subtype and control samples to differentially expressed genes from 5 neuropsychiatric disorders identified in Gandal et al 2018. P-values were calculated using a two-sided Mann-Whitney U test. For boxplots, the center of the box is the median value, the bounds of the box are the 75th and 25th percentile values, the whiskers extend out from the box to 1.5 times the interquartile range of the box, and outlier values are presented as individual points. ASD: autism spectrum disorder, SCZ: schizophrenia, BD: bipolar disorder, MDD: major depressive disorder, AAD: alcoholism.
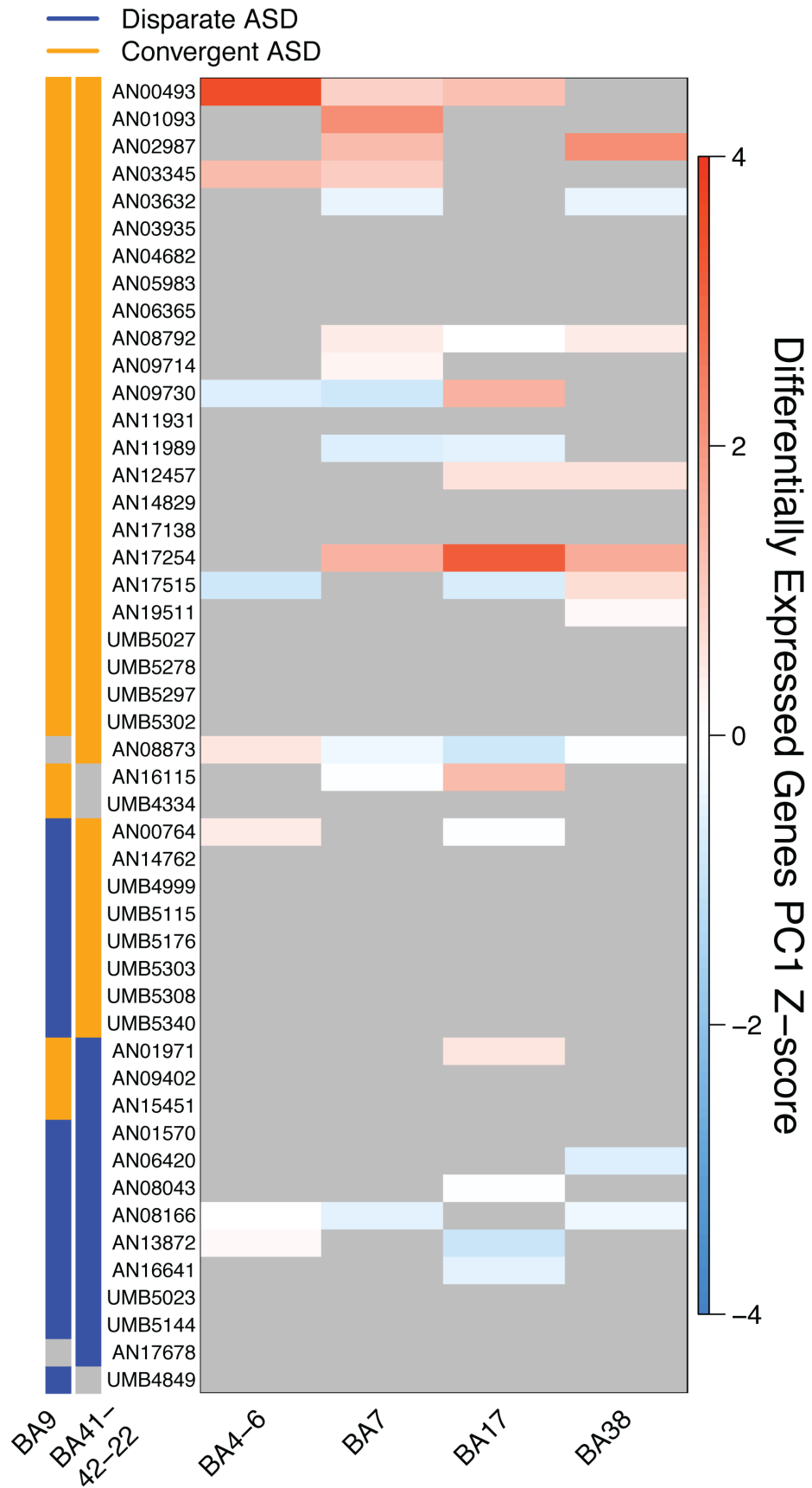
Supplementary Figure 15. Differential gene expression signature of ASD individuals in additional cortical regions. Loadings on the first principal component of differentially expressed genes from Parikshak et al 2016 in four additional cortical regions from Gandal et al 2018. Regions not assessed in a particular individual are colored in grey.