Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

LINE-1 and *Alu* methylation Signatures in Autism Spectrum Disorder and Their Function in The Regulation of Autism-Related Genes

Thanit Saeliw
Chulalongkorn University Faculty of Allied Health Sciences
Tiravut Permpoon
Mahidol University Faculty of Medicine Siriraj Hospital
Nutta Iadsee
Mahidol University Faculty of Medicine Siriraj Hospital
Tewin Tencomnao
Chulalongkorn University Faculty of Allied Health Sciences
Tewarit Sarachana
Chulalongkorn University Faculty of Allied Health Sciences
Daniel Green
University of Liverpool Faculty of Health and Life Sciences
Chanachai Sae-Lee (chanachai.sae@mahidol.ac.th)
Mahidol University, Faculty of Medicine, Siriraj Hospital https://orcid.org/0000-0001-7758-7653

Research Article

Keywords: Autism Spectrum Disorder, DNA methylation, LINE-1, Alu element, Epigenetic age

Posted Date: December 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1194570/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Abstract

Background

Long interspersed nucleotide element-1 (LINE-1) and *Alu* elements are retrotransposons whose abilities cause abnormal gene expression and genomic instability. Several studies have focused on DNA methylation profiling of gene regions, but the locus-specific methylation of LINE-1 and *Alu* elements has not been identified in autism spectrum disorder (ASD).

Methods

Here, DNA methylation age was predicted using Horvath's method. We interrogated locus- and family-specific methylation profiles of LINE-1 and *Alu* elements (22,352 loci) in ASD blood using publicly-available Illumina Infinium 450K methylation datasets from heterogeneous ASD (n = 52), ASD with 16p11.2 del (n = 7), and ASD with *Chromodomain Helicase DNA-binding 8* (*CHD8*) variants (n = 15). The differentially methylated positions of LINE-1 and *Alu* elements corresponding to genes were combined with transcriptome data from multiple ASD studies. ROC curve was conducted to examine the specificity of target loci.

Results

Epigenetic age acceleration was significantly decelerated in ASD children over the age of 11 years. DNA methylation profiling revealed LINE-1 and *Alu* methylation signatures in each ASD risk loci by which global methylation were notably hypomethylated exclusively in ASD with *CHD8* variants. When LINE-1 and *Alu* elements were clustered into subfamilies, we found methylation changes in a family-specific manner in L1P, L1H, HAL, *AluJ*, and *AluS* families in the heterogeneous ASD and ASD with *CHD8* variants. Our results showed that LINE-1 and *Alu* methylation within target genes is inversely related to the expression level in each ASD variant. Moreover, LINE-1 and *Alu* methylation signatures can be used to predict ASD individuals from non-ASD.

Limitations

Integration of methylome and transcriptome datasets was performed from different ASD cohorts. The small sample size of the validation cohort used post-mortem brain tissues and necessitates future validation in a larger cohort.

Conclusions

The DNA methylation signatures of the LINE-1 and *Alu* elements in ASD, as well as their functional impact on ASD-related genes, have been studied. These findings are considered for further research into DNA methylation profiles and the expression of the LINE-1 and *Alu* elements in post-mortem brain tissue, which has been linked to ASD pathogenesis.

Background

Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder characterized by two behavioral impairments: (i) deficits in social interactions and communication, and (ii) restricted interests and repetitive behaviors, according to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition criteria (1). According to the Centers for Disease Control and Prevention (CDC), the prevalence of ASD has risen dramatically over the last decade due to better screening methods (2). In 2016, ASD affected approximately one out of every 54 children in the United States (2). ASD is currently understood as a multifactorial disorder, with the precise causes remaining unknown. Over the last two decades, research has attempted to elucidate the genetic origin of the disorder. However, genetic aberration is only found in 10–20% of ASD cases. In total, more than 60% of people with ASD are idiopathic (3). Several studies have shown that ASD clinical phenotypic heterogeneity is influenced by a combination of genetic and environmental factors (4–6). This evidence has highlighted non-genetic factors such as epigenetics (DNA methylation (DNAm)) and environmental interactions as key players in ASD progression. Additionally, some genetic factors that increase the risk of ASD, but only a few loci have a high impact on ASD (7). The

16p11.2 deletion (16p11.2 del) and *Chromodomain helicase DNA-binding domain 8* (*CHD8*) variants are high genetic risk factors for ASD (8, 9). People with 16p11.2del are usually characterized by developmental delay, intellectual disability, or ASD (9). *CHD8* is strongly associated with ASD and other neurodevelopmental disorders including schizophrenia and intellectual disability (2, 10).

Epigenetics is a family of heritable mechanisms that elicit control of gene expression without modification to DNA sequences (11). Examples of epigenetic mechanisms are DNAm, RNA modification, and histone modifications (12). DNAm, the most frequently studied epigenetic modification, involves the addition of methyl groups to DNA. Depending on its genomic location, the addition of a methyl groups to the 5th carbon atom of cytosine can have repressive or inductive effects on the gene expression. When DNAm is not properly maintained or established, methylation abnormalities can manifest in disease development. DNAm patterns are well known to show tissue-specific differentially methylated regions (DMRs). However, most loci present similar DNAm levels across a wide variety of tissue types. Interestingly, recent work has demonstrated the utility of blood as a surrogate for human brain tissue CpG methylation (13). Therefore, blood-based epigenetic biomarkers have the potential to serve as non-invasive biomarkers for otherwise inaccessible tissues. For instance, some epigenetic markers in blood have been identified as biomarkers in early stages of Alzheimer's disease (14). Similar findings have also arisen in ASD, a recent meta-analysis of blood-based DNA demonstrated evidence of the associations between blood-based and brain samples in comparison between ASD and controls (15).

Long interspersed nucleotide element-1 (LINE-1) and *Alu* elements are known as non-long terminal repeat retrotransposons that can replicate and insert themselves into different locations within the host genome. LINE-1 and *Alu* elements make up more than 25% of the human genome and have a copy number of over one million elements (16). These repetitive elements (REs) can affect the expression of host or neighboring protein-coding genes through introducing alternative promoters or enhancers, novel splicing sites, and epigenetic alteration through DNAm (16). Subfamilies of LINE-1 and *Alu* elements can be subcategorized by identifying variants in their sequences that have accumulated in the evolutionary heritage (17, 18). LINE-1 has been classified into three main subfamilies during early primate evolution including L1M (mammalian-specific, oldest), L1P (primate-specific, intermediate), and L1H (human-specific, youngest) subfamilies (19). *Alu* elements have been classified into three main subfamilies including *AluJ* (oldest), *AluS* (intermediate), and *AluY* (youngest) (17). The ability to transposition has been lost in the oldest subfamilies of both LINE-1 and *Alu*, whereas the intermediate and young subfamilies (L1PA, L1H, *AluS*, and *AluY*) are active and capable of jumping (20).

Current evidence suggests that aberrant DNAm of LINE-1 and *Alu* elements links to several diseases: ASD (21), pre-symptomatic dementia in type 2 diabetes (22), and chronic lymphocytic leukemia (23). Whole-genome sequencing investigation of the brains of individuals with ASD revealed that LINE-1 and *Alu* elements have a larger number of insertions than in normal brain tissues (24). The binding of methyl-CpG binding protein 2 (MeCp2), transcriptional repressor, to the LINE-1 promoter was dramatically reduced, and this was related to LINE-1 overexpression in ASD brains (25). The functional impact of LINE-1 and *Alu* elements in the ASD is currently unknown. One possibility is that LINE-1 and *Alu* elements act as enhancers or alternative promoters for host genes. Our recent study discovered a link between LINE-1/*Alu* elements and gene expression in blood transcriptome, implying that LINE-1 and *Alu* may influence the expression of host genes in ASD (21, 26). Additionally, we also found changes in global methylation of LINE-1 and *Alu* elements in the lymphoblastoid cell line of the ASD subgroup based on clinical phenotypes (21, 26). According to a recent study using blood samples from ASD, one of CpG sites within the LINE-1 sequence showed a slight decrease of methylation levels in ASD compared to unaffected controls but its methylation level was highly significant in ASD with mental regression (27). However, locus- and family-specific methylation patterns of LINE-1 and *Alu* elements in ASD blood have not been reported.

Here, we intended to investigate the DNAm profile of LINE-1 and *Alu* elements, as well as the functional impact on regulation genes located nearby these elements. Using Illumina Infinium 450K annotation, CpG sites mapping to LINE-1 and *Alu* families were identified from DNAm data (GSE113967 and GSE131706) obtained from the NCBI Gene Expression Omnibus (GEO) database. We calculated DNAm Age in ASD blood using Horvath's method. Differential methylation of LINE-1 and *Alu* elements was examined in a global, locus- and subfamily-specific manner for each ASD variant, including heterogeneous ASD

(n = 52), ASD with 16p11.2 del (n = 7), and ASD with *CHD8* variants (n = 15). Biological functions and interactome networks of genes located nearby LINE-1 and *Alu* elements were predicted by ingenuity pathway analysis (IPA). We subsequently identified these genes that were reproducibly differentially expressed in transcriptome data obtained from multiple ASD cohorts. Finally, we used logistic regression to assess the accuracy of a unique LINE-1 and *Alu* methylation signatures in discriminating ASD with a genetic variant from controls.

Methods

Data collection

Differentially methylated retrotransposon loci were identified in publicly available Illumina Infinium 450K datasets through GEO DataSets: http://www.ncbi.nlm.nih.gov/gds (28): GSE113967 (29). In this dataset, ethical approval was granted by the Research Ethics Boards of the respective institutions (University of Michigan SickKids, Holland Bloorview Kids Rehabilitation Hospital, Western University, McMaster University) (29). Data were collected from the heterogeneous ASD (n = 52), ASD with confirmed typical 600 Kbp deletion in 16p11.2 del (n = 7), ASD with confirmed de novo *CHD8* sequence variants (n = 15), and age-matched controls (non-ASD) (n = 48) (Table 1). Validation was performed in a cohort of genome-wide DNAm profiling of post-mortem brain tissue in the subventricular zone of the lateral ventricles from five individuals with ASD and five without (GSE131706) (30) (Table 1).

Genes with differently methylated loci were analyzed in publicly available gene expression datasets from publicly available datasets accessed via GEO using the following inclusion criteria: (1) the study must include ASD cases and controls; (2) the study must use microarray/RNA-seq technology; and (3) the study must use blood or post-mortem brain tissues. Finally, we obtained seven ASD studies, four of which used blood and three of which used post-mortem brain tissues (see Additional file 1).

Table 1
Characteristics of the Infinium450K datasets in the present study including ASD and non-ASD. The numbers of patients for each group are given for each gender. The ages for each group are given as means ± standard deviation.

Groups	Gender		Age
	Male	Female	
GSE113967: Blood			
ASD	45	7	10.45 ± 3.31
Non-ASD	33	15	10.22 ± 3.95
ASD+*CHD8*	11	4	8.97 ± 4.11
ASD+16p11.2 del	6	1	6.78 ± 4.05
GSE131706: Subventricular zone of the lateral ventricles			
ASD	5	-	7.07 ± 1.78
Non-ASD	5	-	7.70 ± 1.40

Epigenetic clock

DNAm Age was predicted using Horvath's method, which is available at methylclock package through the DNAmAge function (31). Horvath's method employed 353 CpG sites with corresponding coefficient values to define DNAm Age as predicted age,

in years (32). Age acceleration residuals were obtained from a multivariate model regression of DNAm Age on chronological age and estimates of blood cell counts. Age acceleration residuals were assessed in all ASD patients by dividing individuals into two groups of ≤ 10 years old and ≥ 11 years old, based on the onset of adolescence (33).

Differential methylation of retrotransposon subfamilies

Methylation datasets were normalized using the single-sample normalized (ssNoob) method in minfi package (34). CpG sites within retrotransposons were identified using Illumina Infinium 450K annotation (23). To identify the variant-associated differential methylation of REs, probes with single nucleotide polymorphisms (SNPs) located at or within 10 base pairs of the target CpG site were included in the analysis. The CpG sites were mapped to LINE-1, *Alu*, half-L1 (HAL1), fossil *Alu* monomer (FAM), free right *Alu* monomer (FRAM), and free left *Alu* monomer (FLAM). Due to the evolution age of the REs, LINE-1 elements were clustered into oldest (L1M, mammalian-wide), intermediate (L1P, primate-specific), and youngest (L1HS, human-specific and L1PA, primate-amplified). Concomitantly, *Alu* elements were categorized into *AluJ* (oldest), *AluS* (intermediate) and *AluY* (youngest).

Mean β value across all loci of REs was calculated as global DNAm in non-ASD, ASD with 16p11.2 del, and ASD with *CHD8* variants. Differential methylation of LINE-1 and *Alu* subfamilies between 1) non-ASD vs ASD, 2) non-ASD vs ASD with 16p11.2 del, and 3) non-ASD vs ASD with *CHD8* variants, were identified. Differentially methylated positions (DMPs) to ASD were examined in the validation dataset. DMPs were identified in ASD, ASD with 16.p11.2 del, ASD with *CHD8* variants and non-ASD with 16.p11.2 del, by two-tailed t-test with correction for false discovery rate (FDR) using the Benjamini-Hochberg (BH) method (35) and significance defined as $P_{FDR} \leq 0.05$. To find the unique DMPs of each data set, the significant loci from 1) to 3) comparisons were computed to create Venn diagrams (https://bioinfogp.cnb.csic.es/tools/venny/).

Differential gene expression analysis

The expression data of ASD studies were obtained from the GEO DataSets. The data from each study were analyzed separately using the Multiple Experiment Viewer (MeV) program (microarray software suite) (36). Firstly, the data were filtered using a 70% cut-off filter to remove probes that were missing in > 30% of samples. The available data were then used for the identification of differentially expressed genes (DEGs) in ASD vs non-ASD cohort by using the Significance Analysis of Microarrays (SAM). The FDR and q-value less than 5% were considered as significantly DEGs.

RNA-sequencing (RNA-seq) data were obtained from the Sequence Read Archive database and re-analyzed using the Galaxy platform (https://usegalaxy.org/) (37). The quality control of RNA-seq data was assessed by *fastp* tool (38). The cleaned reads were then mapped to the human reference genome (GRCh38/hg38) using HISAT2 (39) and quantified using the Subread package FeatureCounts (40). Differential expression analysis was performed using the DESeq2 package (41). The read counts were normalized using the median ratio method of the DESeq2 and the remove unwanted variation (RUV) tool (39). The genes with a *p*-value (*p*) with Benjamini–Hochberg correction of less than 0.05 were considered significant.

Gene functions and pathway analysis

To predict biological functions and gene regulatory networks associated with LINE-1 and *Alu* elements, a list of genes located nearby DMPs of LINE-1 and *Alu* elements for each ASD variant were submitted to the Ingenuity Pathway Analysis software (IPA: QIAGEN Inc.,https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis) (42). Gene regulatory networks were highlighted with $\log_2$ fold change of DNAm level.

Identification of target loci in ASD with each genetic variant

The target RE loci of ASD with each genetic variant were identified by taking the unique DMPs from the Venn diagrams and re-analyzing the different methylation of RE loci in ASD vs ASD with 16.p11.2 del or with *CHD8* variants. We only selected loci which were significant in all three conditions (non-ASD vs ASD, ASD vs ASD with 16.p11.2 del and ASD vs ASD with *CHD8* variants) by two-tailed t-test with correction for multiple hypothesis testing using the BH method and significance defined as

$P_{FDR} \leq 0.05$. Moreover, the DMRs, located nearby the significantly distinct DMPs, were identified in ASD with 16.p11.2 del or with *CHD8* variants.

Statistical analyses

Differentially methylated loci were identified by two-tailed t-tests with significance defined as $P_{FDR} \leq 0.05$ by the BH method or $p \leq 0.05$ for validation data due to a low power in the analysis. Two-tailed t-tests were used to identify significant changes in age acceleration residuals among non-ASD and ASD. Pearson's correlation coefficient was used to compare DNAm Age and chronological age. Fisher's exact test was used to identify enrichment by genomic location of REs. DEGs were identified using SAM analysis with significance defined as FDR ≤ 0.05 by the BH method. Gene function and pathway analysis were performed in IPA using Fisher's exact test with BH correction for multiple testing ($P_{FDR} \leq 0.05$ was considered to be significant). All statistical analyses were performed in R (version 4.0.5) and RStudio (version 1.4.1103) using the ggplot2, plotROC, pheatmap, and GraphPad Prism (version 7.0b); data are presented as mean ± SD, and $p \leq 0.05$ were considered to be significant.

Results

Epigenetic age acceleration (EAA) changes in ASD blood

DNAm Age was predicted using Horvath's method, which employs 353 CpG sites (32). EAA was then estimated as the residual resulting from a multivariate model regressing the DNAm Age estimate on chronological age. Firstly, we correlated chronological age and DNAm Age in non-ASD (n = 59, combining non-ASD and non-ASD with 16.p11.2 del) and ASD (n = 72, combining ASD, ASD with 16p11.2 del, and ASD with *CHD8* variants). The results show that DNAm Age was significantly correlated with chronological age in non-ASD (r = 0.72, $p < 0.0001$), but the correlation was lower in the ASD cohort (r = 0.49, $p < 0.0001$) (Figure 1A and 1B). The expected correlation found in the non-ASD group indicated that DNAm Age in our analyses can be determined EAA. The chronological age difference between the two cohorts was not significant (ASD: 9.94, non-ASD: 9.68, $p = 0.723$) (Figure 1C), while the DNAm Age of the ASD cohort was considerably decreased (ASD: -0.84, non-ASD: 0.84, $p = 0.00078$) (Figure 1D). Next, we classified non-ASD and ASD cohorts by ≤ 10 years old and ≥ 11 years old. As for previous comparisons, chronological age did not differ between ASD and non-ASD in both groups (Figure 1E). In the ASD cohort older than 11 years old, EAA shows a significant deceleration compared with an age-matched non-ASD cohort but this deceleration was not found in ASD younger than 11 years old (Figure 1F). These findings suggest that people with ASD were epigenetically younger than their chronological age which may be occurring in response to various environmental risk factors.

DNAm profile of LINE-1 and *Alu* elements in the heterogeneous ASD

A total of 22,352 probes mapping to LINE-1 and *Alu* elements were identified on the Infinium 450K platform for differential DNAm analysis. The analyses were performed for heterogeneous ASD, ASD with 16p11.2del, and ASD with *CHD8* variants, versus non-ASD. Firstly, we measured the global methylation by combining all positions mapping to LINE-1 and *Alu* elements as the total of CpGs. In the comparison of heterogeneous ASD against non-ASD (Figure 2A), there was no significant difference in global methylation between these cohorts ($\Delta\beta = 0.003$, $p = 0.098$). However, when we performed the methylation profile of REs by which LINE-1 and *Alu* positions were analyzed separately, we found that 2,802 (LINE-1) and 4,363 (*Alu*) DMPs were significantly differentially methylated ($P_{FDR} < 0.05$) in the heterogeneous ASD compared to non-ASD (Figure 2B and 2C, Figure 3). All these loci included 2,471 hypomethylated loci (LINE-1: 1,437 loci, *Alu*: 1,304 loci; $P_{FDR} < 0.05$) and 4,424 hypermethylated loci (LINE-1:1,365 loci, *Alu*: 3,059 loci; $P_{FDR} < 0.05$). Due to the different activity of subfamilies of RE, LINE-1 and *Alu* elements were clustered by evolution age into three categories including old age (L1M, *AluJ*), intermediate age (L1P, L1PB, *AluS*), young age (L1HS, L1PA, *AluY*), and related (HAL1, FAM, FLAM, FRAM). The methylation of LINE-1 and *Alu* elements were changed in a subfamily-specific manner. We discovered that LINE-1 was considerably hypermethylated in young and intermediate age families, including L1H ($\Delta\beta = 0.013$, $p = 0.00001$) and L1P ($\Delta\beta = 0.005$, $p = 0.027$), but HAL1 was hypomethylated ($\Delta\beta = -0.003$, $p = 0.03$) (Figure 2D). Methylation of *Alu* elements was significantly hypermethylated in the old

and intermediate age families: *AluJ* (Δβ = 0.006, p = 0.004) and *AluS* (*Δ*β = 0.005, p = 0.016), respectively (Figure 2E). These findings indicated that methylation of LINE-1 and *Alu* elements in the heterogeneous ASD was altered in family- and locus-specific manner rather than globally.

LINE-1 and *Alu* methylation signatures in the homogeneous ASD (16p11.2del and *CHD8* variants)

Due to the heterogeneity in the ASD population, we also intended to investigate the methylation profile of LINE-1 and *Alu* elements in genetically homogeneous ASD, as identified in the original article of GSE113967, including ASD individuals with 16p11.2del (n = 7) and *CHD8* variant (n = 15). As for the results of ASD with 16p11.2del compared with non-ASD, we found no significant changes in the global methylation compared with non-ASD (Δβ = -0.002, p = 0.771, Figure 4A). However, the analyses identified 70 significantly locus-specific DMPs of REs in ASD with 16p11.2del including 27 DMPs at LINE-1 (5 hypomethylated loci, 22 hypermethylated loci, Figure 4B) and 43 DMPs at *Alu* elements (23 hypomethylated loci, 20 hypermethylated loci, Figure 4C). When LINE-1 and *Alu* positions were categorized into families, there was no significant difference in methylation of LINE-1 and *Alu* elements by family (Figure 4D and 4E).

Subsequently, we analyzed data for ASD with *CHD8* variants by using the same approach. We found that global methylation was exclusively hypomethylated in the ASD with *CHD8* variants (Δβ = -0.006, p = 0.042, Figure 5A). Analyzing by the position, the majority of DMPs at LINE-1 and *Alu* elements were hypomethylated (616 loci or 88.63%, P_{FDR}<0.05) of the total identified 695 DMPs observed (Figure 5B and 5C). Among all significant DMPs in the ASD with *CHD8* variants, 528 DMPs were mapped to LINE-1, while 167 DMPs were *Alu* elements. Moreover, changes in LINE-1 and *Alu* methylation regarding to their families were observed in ASD with *CHD8* variants. In contrast to the differences found in the heterogeneous ASD, young and intermediate age LINE-1 families were significantly hypomethylated including L1H (Δβ = -0.015, p = 0.0038) and L1P (Δβ = -0.010, p = 0.0186) (Figure 5D). Hypomethylation of *Alu* elements was also observed in old age, intermediated age, and related families: *AluJ* (Δβ = -0.010, p = 0.0154), *AluS* (Δβ = -0.008, p = 0.0443), FAM (Δβ = -0.008, p = 0.0399), FRAM (Δβ = -0.007, p = 0.02551) respectively (Figure 5E). These findings suggest that DNAm signatures were a widespread reduction in LINE-1 and *Alu* regions which occurred at a specific family in the ASD with *CHD8* variants but not in ASD with 16p11.2del.

Genomic distribution of LINE-1 and *Alu* methylation in heterogenous and homogenous ASD.

To determine the differential DNAm of LINE-1 and *Alu* elements by genomic features, we performed enrichment analysis using Fisher's exact test. CpG positions at LINE-1 and *Alu* elements were categorized to 1500 and 200 within the transcriptional start site (TSS1500 and TSS200, respectively), the 5' untranslated region (5'UTR), the first exon (1st exon), gene body (Body), and 3' untranslated region (3'UTR). In the heterogeneous ASD signatures, CpG sites at LINE-1 were significantly enriched in TSS1500 (p = 0.0005) and Body (p < 0.0001) (see Additional file 2A). Whereas *Alu* elements were significantly enriched in TSS1500 (p < 0.0001), 5'UTR (p < 0.0001), Body (p < 0.0001), and 3'UTR (p = 0.0086) (see Additional file 3A). However, DNAm across all retrotransposons by genomic location did not significantly differ between non-ASD and heterogenous ASD (see Additional file 2B and 3B). DNAm signatures of the ASD with 16p11.2del and *CHD8* variants were significantly enriched in Body (p = 0.04) and TSS1500 (p < 0.0001) respectively (see Additional file 4: Figure S5B, S5C). This result shows that the changes of probes mapping to TSS1500 and gene bodies are more likely to have a functional impact on gene expression in ASD in both heterogenous and homogenous ASD.

Biological functions and pathways of LINE-1 and *Alu* methylation signatures in ASD and ASD variants.

To determine the biological significance of LINE-1 and *Alu* methylation signatures identified in each ASD cohort, we predicted the biological function and pathway of genes located nearby DMPs of LINE-1 and *Alu* elements using IPA software. We found that neurological diseases were significantly enriched among genes associated with LINE-1 and *Alu* methylation signatures in the heterogeneous ASD (p range: 0.00495 - 3.33E-26, 2274 genes) and ASD with *CHD8* variants (p range: 0.0258 – 0.000117, 302 genes) as shown in Additional file 5 and 6. The categories ASD and intellectual disability were exclusively associated with LINE-1 and *Alu* methylation signatures in the heterogeneous ASD (p = 2.56E-06, 253 genes). Whereas Huntington's disease, familial encephalopathy, and brain lesion were commonly associated with both ASD signatures. For ASD with

16p11.2del variant, LINE-1 and *Alu* methylation signatures in this cohort were significantly associated with developmental disorders (p range: 0.0393 – 0.00222, 9 genes) (see Additional file 7). However, only one gene was associated with the disease, possibly caused by a small number of genes associated with LINE-1 and *Alu* methylation of this ASD variant. Additionally, we discovered that several canonical pathways linked to ASD were associated with genes located nearby LINE-1 and *Alu* methylation signatures in each ASD cohort. More precisely, we found that the α-adrenergic signaling pathway was significantly associated in the heterogeneous ASD (p = 0.00269, 28 genes) and ASD with *CHD8* variants (p = 0.00646, 7 genes). Axonal guidance signaling pathway involved in nervous system development was significantly associated with LINE-1 and *Alu* methylation signatures of ASD with 16p11.2del and *CHD8* variants. These results indicate that genes associated with LINE-1 and *Alu* methylation signatures in ASD were involved with neurological diseases and ASD-comorbid disorders as well as canonical pathways known to be implicated in ASD. The list of all significant biological functions and pathways in each ASD variant is shown in Additional file 5-7.

Interactome networks or gene regulatory networks revealed the interaction of genes located nearby LINE-1 and *Alu* methylation signatures of each ASD variant. The functions and pathways implicated in ASD were highlighted in the networks. The interactome of the heterogeneous ASD was associated with ASD and mental retardation, as well as canonical pathways implicated in ASD such as retinoic acid receptor (RAR) and AMP-activated protein kinase (AMPK) signaling (Figure 6). In ASD with 16p11,2del, we found that the interactome related to axonal guidance and sirtuin signaling pathway (see Additional file 8: Figure S8A). The interactome of ASD with *CHD8* was related to familial encephalopathy and movement disorder which conditions found in ASD individuals (43, 44). The interactomes were also associated with neuronal function including axonal guidance and synaptogenesis signaling pathways (see Additional file 8: Figure S8B and S8C).

Identification of unique target loci located nearby LINE-1 and *Alu* signatures in heterogeneous ASD

To investigate the functional impact of locus-specific LINE-1 and *Alu* methylation to target gene or neighboring gene expression in the ASD, we identified DEGs from multiple ASD studies obtained from the GEO DataSets. This approach reflected the heterogeneity of the ASD population because these studies were compiled from a different ASD cohort. There were 12,419 DEGs identified from seven datasets including four studies (one study used peripheral blood samples and three studies used post-mortem brain tissues from ASD individuals) (see Additional file 1). We subsequently overlapped the list of DEGs with differentially methylated genes (DMGs: genes located nearby LINE-1 and *Alu* signatures). The overlapping revealed 1,847 DMGs in the heterogeneous ASD that were differentially expressed in several ASD studies, with 155 of them being autism-related genes in the SFARI database. We identified 43 top DMGs, $|\Delta\beta| \geq 5\%$, inversely related to gene expression, and differentially expressed in at least two studies (Table 2). Interestingly, two of the top DMGs, potassium voltage-gated channel subfamily Q member 3 (*KCNQ3*) and ubiquitin conjugating enzyme E2 H (*UBE2H*), were genes in the SFARI database and were enriched in the gene regulatory network related to ASD and mental retardation (Figure 6).

The genomic regions of LINE-1 and *Alu* methylation signatures within the DMGs are shown in Figure 7. We identified DMRs by mapping all probes located nearby LINE-1 and *Alu* signatures using the UCSC genome browser. The findings revealed that *AluSg7* (cg16926147), which is located on the gene body of the *KCNQ3* gene (Figure 7A), was hypermethylated and *KCNQ3* expression level was significantly reduced in blood and post-mortem brain tissues. Interestingly, we discovered that several probes in this region, including those in the promoter region were not changed. This result suggests that LINE-1 and *Alu* methylation at DMRs may facilitate gene expression indicated by the inverse relationship between LINE-1/*Alu* methylation and gene expression. As well as *AluY* (cg08998414) within *UBE2H* gene (Figure 7B) and L1PA3 (cg24094412) within hyperpolarization activated cyclic nucleotide gated potassium channel 1 (*HCN1*) (Figure 7C), we also observed that *AluY* and L1PA3 methylation were inversely related to the gene expression levels in both blood and brain tissues of ASD cohort. Moreover, we found several DMGs that were not reported in the SFARI database but the expression of these DMGs in the blood and post-mortem brain tissues was inversely related to LINE-1 and *Alu* methylation such as *N-deacetylase and N-sulfotransferase 1* (*NDST1*) (cg12611243: L1MC1), *ubiquitin specific peptidase 6* (*USP6*) (cg23416909: L1M5), and *formin binding protein 1* (*FNBP1*) (cg13916261: *AluSg*). These results suggest that DMPs at LINE-1 and *Alu* elements may affect the expression of genes located nearby these DMPs in the heterogeneous ASD cohort.

Table 2
Utilized gene expression microarrays/RNA-sequencing for the differential gene expression analysis of the target genes of heterogeneous ASD.

Methylome data				**Transcriptome data**				
Probe ID	Elements	Delta	P_{FDR}	GSE	Gene ID	Gene	log2FC	q-value
cg02571470	L1MC5	0.061	1.61E-05	GSE18123	1563708_at	*SFXN5*	-0.348	0.0164
cg02571470	L1MC5	0.061	1.61E-05	GSE18123	241999_at	*SFXN5*	-0.407	0.0386
cg21314304	AluJb	0.054	2.68E-05	GSE25507	226298_at	*RUNDC1*	-0.141	0.0272
cg21314304	*AluJb*	0.054	2.68E-05	GSE28521_FC	ILMN_1733875	*RUNDC1*	-0.234	0.0319
cg21314304	*AluJb*	0.054	2.68E-05	GSE59288	146923	*RUNDC1*	-0.452	0.0001
cg04668642	*AluSx*	0.051	2.68E-05	GSE64018	ENSG00000166780	*C16orf45*	-0.224	0.0421
cg04668642	*AluSx*	0.051	2.68E-05	GSE28521_FC	ILMN_1687821	*C16orf45*	-0.256	0.0287
cg02616069	*AluJb*	0.069	3.02E-05	GSE18123	215584_at	*HECW1*	-0.259	0.0068
cg02616069	*AluJb*	0.069	3.02E-05	GSE59288	23072	*HECW1*	-0.805	<0.0001
cg15531814	*AluJo*	0.059	4.42E-05	GSE89594	A_23_P310257	*KLK2*	-0.161	0.0176
cg15531814	*AluJo*	0.059	4.42E-05	GSE18123	1555545_at	*KLK2*	-0.138	0.0254
cg11204311	*AluJr*	0.060	4.54E-05	GSE28521_FC	ILMN_1679796	*TOMM20*	-0.506	0.0110
cg11204311	*AluJr*	0.060	4.54E-05	GSE59288	9804	*TOMM20*	-0.176	0.0438
cg11204311	*AluJr*	0.060	4.54E-05	GSE64018	ENSG00000173726	*TOMM20*	-0.231	0.0459
cg04027778	*AluSc*	0.054	4.66E-05	GSE28521_TC	ILMN_1712705	*RAB40C*	-0.258	0.0313
cg04027778	*AluSc*	0.054	4.66E-05	GSE18123	227269_s_at	*RAB40C*	-0.429	0.0133
cg04027778	*AluSc*	0.054	4.66E-05	GSE18123	1569396_at	*RAB40C*	-0.309	0.0025
cg04027778	*AluSc*	0.054	4.66E-05	GSE59288	57799	*RAB40C*	-0.322	0.0131
cg06719602	*AluSx1*	0.050	5.57E-05	GSE28521_FC	ILMN_1706238	*CSE1L*	-0.202	0.0384
cg06719602	*AluSx1*	0.050	5.57E-05	GSE28521_FC	ILMN_1665797	*CSE1L*	-0.184	0.0449

Methylome data				Transcriptome data				
cg23935361	*AluJb*	0.066	5.82E-05	GSE42133	ILMN_1738093	*RNFT2*	-0.102	0.0090
cg23935361	*AluJb*	0.066	5.82E-05	GSE59288	84900	*RNFT2*	-0.259	0.0467
cg17429234	*AluSp*	0.061	6.88E-05	GSE59288	1780	*DYNC1I1*	-0.852	<0.0001
cg17429234	*AluSp*	0.061	6.88E-05	GSE28521_FC	ILMN_1690397	*DYNC1I1*	-0.516	0.0081
cg17429234	*AluSp*	0.061	6.88E-05	GSE64018	ENSG00000158560	*DYNC1I1*	-0.324	0.0410
cg23376467	*AluSq*	0.077	7.13E-05	GSE18123	1556907_at	*ZNF474*	-0.292	0.0346
cg23376467	*AluSq*	0.077	7.13E-05	GSE89594	A_33_P3360565	*ZNF474*	-0.205	0.0200
cg06471678	*AluYc*	-0.065	7.85E-05	GSE25507	233694_at	*HSPA1L*	0.111	0.0307
cg06471678	*AluYc*	-0.065	7.85E-05	GSE42133	ILMN_1654566	*HSPA1L*	0.107	0.0300
cg15074424	*AluSx1*	0.083	1.01E-04	GSE28521_FC	ILMN_2212354	*WDR46*	-0.201	0.0364
cg15074424	*AluSx1*	0.083	1.01E-04	GSE28521_TC	ILMN_2212354	*WDR46*	-0.303	0.0335
cg02747612	*AluSq*	-0.068	1.02E-04	GSE42133	ILMN_1802053	*ZNF91*	0.162	0.0158
cg02747612	*AluSq*	-0.068	1.02E-04	GSE18123	236128_at	*ZNF91*	0.306	0.0474
cg13916261	*AluSg*	-0.100	1.05E-04	GSE59288	23048	*FNBP1*	0.354	0.0128
cg13916261	*AluSg*	-0.100	1.05E-04	GSE42133	ILMN_1797342	*FNBP1*	0.156	<0.0001
cg02827046	L1PA16	0.099	1.17E-04	GSE89594	A_33_P3420900	*PATE2*	-0.221	0.0142
cg02827046	L1PA16	0.099	1.17E-04	GSE42133	ILMN_2133784	*PATE2*	-0.157	0.0236
cg07930329	*AluSx3*	0.056	1.21E-04	GSE59288	10055	*SAE1*	-0.295	0.0064
cg07930329	*AluSx3*	0.056	1.21E-04	GSE28521_TC	ILMN_1657204	*SAE1*	-0.194	0.0417
cg07930329	*AluSx3*	0.056	1.21E-04	GSE64018	ENSG00000142230	*SAE1*	-0.194	0.0395
cg16926147	*AluSg7*	0.062	1.31E-04	GSE18123	206573_at	*KCNQ3*	-0.093	0.0182
cg16926147	*AluSg7*	0.062	1.31E-04	GSE59288	3786	*KCNQ3*	-0.634	0.0003

Methylome data				Transcriptome data				
cg08499057	*AluJo*	0.057	1.63E-04	GSE28521_FC	ILMN_1781999	*ABCF2*	-0.164	0.0426
cg08499057	*AluJo*	0.057	1.63E-04	GSE28521_FC	ILMN_1669201	*ABCF2*	-0.168	0.0426
cg08499057	*AluJo*	0.057	1.63E-04	GSE28521_TC	ILMN_1669201	*ABCF2*	-0.202	0.0372
cg08499057	*AluJo*	0.057	1.63E-04	GSE18123	207623_at	*ABCF2*	-0.321	0.0053
cg08998414	*AluY*	-0.106	1.85E-04	GSE64018	ENSG00000186591	*UBE2H*	0.130	0.0352
cg08998414	*AluY*	-0.106	1.85E-04	GSE42133	ILMN_1757644	*UBE2H*	0.126	0.0258
cg08998414	*AluY*	-0.106	1.85E-04	GSE25507	222419_x_at	*UBE2H*	0.171	0.0230

Identification of unique target loci located nearby LINE-1 and *Alu* signatures in ASD variants.

To investigate the effects of unique LINE-1 and *Alu* methylation signatures to target gene or neighboring gene expression in the genetically homogeneous ASD, we obtained 39 and 101 DMPs that were found exclusively in the ASD with 16.p11.2 del and *CHD8* variants, respectively (see Additional file 4: Figure S4A). We re-analyzed them for ASD variant versus the heterogenous ASD. Next, we conducted the same strategy used for the heterogeneous ASD to select the candidate DMPs by overlapping with the transcriptome data. The overlapping of unique DMPs with transcriptome data revealed 11 and 31 unique DMGs in the ASD with 16.p11.2 del and *CHD8* variants, respectively (Table 3 and 4). Among the unique DMGs, we found several genes linked to neurodevelopmental disorder and ASD, including *XK related 6* (*XKR6*) (Figure 8A), *zinc finger protein 107* (*ZNF107*) (Figure 8B), and *myeloma-overexpressed gene 2 protein* (*MYEOV2*) (Figure 8C) in the ASD with 16.p11.2 del. The significant DMPs at *AluY* (cg21300361) within *XKR6* was hypermethylated, while as *AluSq* (cg01772945) within *ZNF107* and L1MB3 (cg13749477) within *MYEOV2* were hypomethylated. Interestingly, these genes were differentially expressed in the blood transcriptome of multiple ASD cohorts, and their expression was inversely relative to LINE-1 and *Alu* methylation levels.

For ASD with *CHD8* variants, we found that all LINE-1 and *Alu* elements located on candidate genes were markedly hypomethylated, as expected from global and family-specific methylation levels. These DMPs consist of L1MC5 (cg22706070) within *Euchromatic Histone Lysine Methyltransferase 2* (*EHMT2*) (Figure 9A), *AluJo* (cg06421197) within *caspase 1* (*CASP1*) (Figure 9B), and *AluSx* (cg18699242, cg01963623, cg02169692) within *ubiquitin-specific peptidase 18* (*USP18*) (Figure 9C). *EHMT2* was significantly increased in the blood of ASD, while *CASP1* was increased in both the blood and brain of multiple ASD cohorts (one probe was decreased). These changes were inversely relative to LINE-1 and *Alu* methylation levels within that gene. We found that the expression of *USP18* was not inversely relative to *AluSx* methylation located on the gene. Additionally, the DMRs of *XKR6, ZNF107, MYEOV2, EHMT2*, and *CASP1* genes revealed LINE-1 and *Alu* probes as well as non-LINE-1/*Alu* probes located in the same DMRs (Figure 8A-C and Figure 9A-B).

Table 3
Utilized gene expression microarrays/RNA-sequencing for the differential gene expression analysis of the target genes of ASD with 16p11.2 deletion.

Methylome data				**Transcriptome data**				
Probe ID	Elements	Delta	P_{FDR}	GSE	Gene ID	Gene	log2FC	q-value
cg22062537	L1MB3	-0.079	0.0002	GSE25507	1553515_at	*MYEOV2*	0.107	0.0496
cg13749477	L1MB3	-0.038	0.0004	GSE25507	1553515_at	*MYEOV2*	0.107	0.0496
cg07628769	L1MB3	-0.101	0.0028	GSE25507	1553515_at	*MYEOV2*	0.107	0.0496
cg01772945	*AluSq*	-0.048	0.0003	GSE18123	205739_x_at	*ZNF107*	0.382	0.0017
cg09168728	HAL1	0.042	0.0039	GSE18123	202651_at	*LPGAT1*	0.160	0.0038
cg21300361	*AluY*	0.024	0.0139	GSE18123	1553640_at	*XKR6*	-0.949	0.0206
cg21300361	*AluY*	0.024	0.0139	GSE25507	1553640_at	*XKR6*	0.134	0.0112
cg01270736	*AluJb*	-0.037	0.0167	GSE25507	207289_at	*MMP25*	0.193	<0.0001
cg26620682	L1PA2	-0.058	0.0206	GSE28521_FC	ILMN_1765641	*SEMA3A*	-0.200	0.0343
cg26620682	L1PA2	-0.058	0.0206	GSE59288	10371	*SEMA3A*	-0.590	0.0139
cg05073382	L1MA7	0.153	0.0290	GSE89594	A_23_P258912	*MYOM2*	-1.611	0.0114
cg11438448	*AluSx4*	-0.014	0.0364	GSE18123	233429_at	*SPEF2*	-0.285	0.0484
cg10059324	L1MC4a	0.050	0.0368	GSE59288	8863	*PER3*	0.296	0.0269
cg11859345	*AluSp*	-0.031	0.0474	GSE18123	228766_at	*CD36*	0.171	0.0249
cg11859345	*AluSp*	-0.031	0.0474	GSE18123	206488_s_at	*CD36*	0.273	0.0076
cg11859345	*AluSp*	-0.031	0.0474	GSE25507	209554_at	*CD36*	0.099	0.0131
cg25283432	L1MA8	0.004	0.0482	GSE25507	232421_at	*SCARB1*	0.116	0.0189

Table 4
Utilized gene expression microarrays/RNA-sequencing for the differential gene expression analysis of the target genes of ASD with *CHD8* variants.

Methylome data				**Transcriptome data**				
Probe ID	Elements	Delta	P_{FDR}	GSE	Gene ID	Gene	log2FC	q-value
cg10071848	*AluJb*	-0.033	0.0439	GSE59288	5826	*ABCD4*	0.39	0.0078
cg07730946	*AluJb*	-0.027	0.0128	GSE42133	ILMN_1684585	*ACSL1*	-0.24	0.0344
cg12934569	*AluSx*	-0.043	0.0319	GSE42133	ILMN_1684585	*ACSL1*	-0.24	0.0344
cg11111835	L1MA3	-0.020	0.0359	GSE18123	1553603_s_at	*ATL2*	0.28	0.0248
cg11111835	L1MA3	-0.020	0.0359	GSE18123	237968_at	*ATL2*	-0.25	0.0492
cg04154142	*AluSx1*	-0.063	0.0110	GSE42133	ILMN_1672596	*BCAR1*	-0.09	0.0127
cg04154142	*AluSx1*	-0.063	0.0110	GSE18123	223116_at	*BCAR1*	-0.26	0.0226
cg14115346	L1MB3	-0.038	0.0201	GSE18123	205750_at	*BPHL*	0.28	0.0097
cg10211626	*AluSx1*	-0.059	0.0028	GSE59288	79640	*C22orf46*	0.56	0.0002
cg17129519	*AluSg*	0.097	0.0360	GSE64018	ENSG00000204564	*C6orf136*	-0.18	0.0430
cg17129519	*AluSg*	0.097	0.0360	GSE28521_TC	ILMN_1813236	*C6orf136*	-0.30	<0.0001
cg06421197	*AluJo*	-0.034	0.0266	GSE18123	206011_at	*CASP1*	0.14	0.0122
cg06421197	*AluJo*	-0.034	0.0266	GSE42133	ILMN_2326509	*CASP1*	-0.16	0.0258
cg06421197	*AluJo*	-0.034	0.0266	GSE42133	ILMN_2326512	*CASP1*	-0.17	0.0213
cg06421197	*AluJo*	-0.034	0.0266	GSE42133	ILMN_1727762	*CASP1*	-0.25	<0.0001
cg06421197	*AluJo*	-0.034	0.0266	GSE59288	834	*CASP1*	0.61	0.0096
cg13606720	L1ME3F	-0.044	0.0125	GSE42133	ILMN_3248676	*CBWD3*	-0.11	0.0180
cg11379605	*AluJb*	-0.033	0.0359	GSE59288	965	*CD58*	0.47	0.0159
cg11379605	*AluJb*	-0.033	0.0359	GSE18123	222061_at	*CD58*	0.36	0.0256
cg11379605	*AluJb*	-0.033	0.0359	GSE64018	ENSG00000116815	*CD58*	0.50	0.0328
cg09425611	L1MB7	-0.028	0.0360	GSE18123	209616_s_at	*CES1*	0.44	0.0025
cg00053536	*AluSx*	-0.028	0.0462	GSE25507	206274_s_at	*CROCC*	0.12	0.0458
cg09604414	*AluSx*	-0.067	0.0336	GSE25507	229079_at	*EHMT2*	0.09	0.0471
cg22706070	L1MC5	-0.008	0.0496	GSE25507	229079_at	*EHMT2*	0.09	0.0471
cg14570121	FLAM_A	-0.031	0.0375	GSE59288	2068	*ERCC2*	-0.26	0.0149
cg24803614	*AluSc8*	0.051	0.0341	GSE59288	2495	*FTH1*	0.42	0.0045
cg03085932	L1PA6	-0.043	0.0459	GSE25507	220249_at	*HYAL4*	0.12	0.0179
cg08005007	L1HS	-0.031	0.0095	GSE18123	235111_at	*LSAMP*	-0.23	0.0386
cg08005007	L1HS	-0.031	0.0095	GSE18123	229244_at	*LSAMP*	-0.25	0.0303
cg22993878	L1MB3	-0.021	0.0080	GSE25507	1553515_at	*MYEOV2*	0.11	0.0496
cg03514928	L1PA17	-0.022	0.0470	GSE59288	4897	*NRCAM*	-0.31	0.0104

Methylome data				Transcriptome data				
cg20723844	*AluY*	-0.061	0.0163	GSE42133	ILMN_1789616	*NUPL2*	-0.09	0.0221
cg00984715	*AluSq*	-0.049	0.0224	GSE59288	5465	*PPARA*	0.51	0.0149
cg00984715	*AluSq*	-0.049	0.0224	GSE18123	1560981_a_at	*PPARA*	-0.24	0.0256
cg00984715	*AluSq*	-0.049	0.0224	GSE18123	1558631_at	*PPARA*	-0.46	0.0091
cg24351819	*AluSp*	-0.026	0.0474	GSE59288	55170	*PRMT6*	-0.33	0.0057
cg09521141	*AluSg7*	-0.045	0.0263	GSE28521_FC	ILMN_1765641	*SEMA3A*	-0.20	0.0343
cg09521141	*AluSg7*	-0.045	0.0263	GSE59288	10371	*SEMA3A*	-0.59	0.0139
cg26620682	L1PA2	-0.041	0.0346	GSE28521_FC	ILMN_1765641	*SEMA3A*	-0.20	0.0343
cg26620682	L1PA2	-0.041	0.0346	GSE59288	10371	*SEMA3A*	-0.59	0.0139
cg18221988	L1PA2	-0.043	0.0499	GSE18123	210804_x_at	*SLC8A1*	-0.30	0.0392
cg18221988	L1PA2	-0.043	0.0499	GSE25507	1565306_a_at	*SLC8A1*	0.18	0.0189
cg18221988	L1PA2	-0.043	0.0499	GSE59288	6546	*SLC8A1*	-0.41	0.0245
cg18221988	L1PA2	-0.043	0.0499	GSE18123	235518_at	*SLC8A1*	0.29	0.0057
cg18221988	L1PA2	-0.043	0.0499	GSE64018	ENSG00000183023	*SLC8A1*	-0.26	0.0430
cg07020453	L1MDa	-0.009	0.0457	GSE18123	217968_at	*TSSC1*	-0.21	0.0185
cg08428949	*AluSz*	-0.028	0.0124	GSE64018	ENSG00000128881	*TTBK2*	-0.15	0.0352
cg01810763	*AluSx1*	-0.053	0.0442	GSE64018	ENSG00000198431	*TXNRD1*	0.30	0.0443
cg11191744	*AluY*	-0.050	0.0457	GSE64018	ENSG00000198431	*TXNRD1*	0.30	0.0443
cg02169692	*AluSx*	-0.109	0.0275	GSE42133	ILMN_3240420	*USP18*	-0.31	0.0250
cg02169692	*AluSx*	-0.109	0.0275	GSE42133	ILMN_3240420	*USP18*	-0.31	0.0250
cg18699242	*AluSx*	-0.106	0.0448	GSE42133	ILMN_3240420	*USP18*	-0.31	0.0250
cg18699242	*AluSx*	-0.106	0.0448	GSE42133	ILMN_3240420	*USP18*	-0.31	0.0250
cg15726387	*AluSz6*	-0.044	0.0950	GSE18123	220079_s_at	*USP48*	0.10	0.0053
cg15726387	*AluSz6*	-0.044	0.0950	GSE18123	225925_s_at	*USP48*	0.15	<0.0001
cg06385000	*AluSc*	-0.041	0.0216	GSE18123	227434_at	*WBSCR17*	-0.46	0.0392
cg06385000	*AluSc*	-0.041	0.0216	GSE28521_FC	ILMN_1701557	*WBSCR17*	-0.19	0.0402
cg06385000	*AluSc*	-0.041	0.0216	GSE28521_TC	ILMN_1701557	*WBSCR17*	-0.23	0.0485
cg03870862	L1MB3	-0.063	0.0318	GSE59288	23144	*ZC3H3*	0.27	0.0193

Sensitivity and specificity of unique LINE-1 and *Alu* signatures in ASD variants.

To predict diagnosis of the genetically homogenous ASD by using LINE-1 and *Alu* methylation signatures, we subsequently conducted ROC curves analysis of these loci and other probes within unique DMRs to distinguish each homogenous ASD variant from non-ASD and ASD with non-specific variants. For ASD with 16.p11.2 del, *AluY* within *XKR6* (cg21300361) exhibited high sensitivity and specificity (AUC = 0.905, 95%CI = 0.83-0.98) to distinguish ASD with 16.p11.2 del from non-ASD

and ASD with *CHD8* variants as shown in the ROC curves (Figure 8D). In addition, the ROC curves of *AluSq* within *ZNF107* (cg01772945) and L1MB3 within *MYEOV2* (cg13749477) also exhibited high AUC value (*AluSq*: AUC = 0.900, 95%CI = 0.83-0.97 and L1MB3: AUC = 0.841, 95%CI = 0.74-0.95) (Figure 8E and 8F). In the ASD with *CHD8* variants, LINE-1 and *Alu* methylation signatures within candidate DMGs showed moderate sensitivity and specificity as demonstrated by AUC values (AUC range: 0.712-0.819) compared with the specificity of unique loci in ASD with 16.p11.2 del, including L1MC5 (cg22706070) within *EHMT2* (Figure 9D), *AluJo* (cg06421197) within *CASP1* (Figure 9E), and *AluSx* (cg18699242, cg01963623, cg02169692) within *USP18* (Figure 9F). Our findings suggest that these novel DMPs at the LINE-1 and *Alu* elements could be used for clinical purposes. However, an independent cohort is required for validation, as we were limited by the percentage of ASD individuals affected by these genetic variants.

Discussion

Epigenetic modification is an important mechanism linking environmental and genetic factors, especially during the developmental process. Epigenetic age is a biomarker for age-related epigenetic changes as well as disease-specific (45), but it is unknown whether it contributes to ASD pathogenesis. DNAm age acceleration rates were significantly increased in the blood of mid-childhood ASD (5-11 years) using the method established by Wu et al (46). In this study, DNAm Age was calculated using Horvath's epigenetic clock algorithm, which shows a high correlation between the chronological age and DNAm Age. Interestingly, we discovered DNAm Age in ASD older than 11 years old was significantly decelerated but not in younger patients aged <11 years in our study. This observation is assumed to be related to the external exposures accumulated during childhood leading to developmental delay. Concomitantly, the mean of age acceleration difference in adolescence ASD (12-18 years old) was lower than their unaffected siblings (not statistically significant) (46). Additionally, we looked into the LINE-1 and *Alu* methylation profiles for ASD with these groups. Although DNAm Age was reduced in ASD individuals over the age of 11, the DNAm profile of the LINE-1 and *Alu* elements was significantly altered only in ASD under the age of 11 (see Additional file 9). These findings suggest that the DNAm signature of LINE-1 and *Alu* elements may change early which corresponds to early developmental and behavioral delays or deviancies. The epigenetic delays were also observed in post-mortem brain tissues of children with ASD (aged 5-8 years) (30).

In the heterogeneous ASD, we found no difference in global methylation (LINE-1 and *Alu*) (Figure 2A), which is consistent with our previous studies using lymphoblastoid cell lines that found no difference when all ASD were combined. In addition to Shpyleva's study, total methylation of LINE-1 in the ASD brain was also not significantly altered (25). The possibility is that the aberration of the global methylation levels (LINE-1 and *Alu*) may rely on family-specific REs or restrict to specific locations. Reducing the heterogeneity of ASD by classifying ASD based on clinical phenotype may be beneficial, as demonstrated by previous findings from our investigators (21, 26, 47–49). Subcategorizing ASD allowed us to observe the hypomethylation of global methylation in ASD with *CHD8* variants.

The aberration of LINE-1 or *Alu* elements during development may cause double-strand DNA breaks and DNA damage leading to the process of neurodegeneration (50, 51). Furthermore, identification of LINE-1 and *Alu* subfamilies has led to a better understanding of the association between the REs and human diseases because some subfamilies of LINE-1 and *Alu* elements remain active (20). To the best of our knowledge, our study is the first to identify the locus-specific methylation at LINE-1 and *Alu* elements in a subfamily-specific manner of the ASD blood samples. In this study, 7,165 DMPs at LINE-1 and *Alu* elements were identified in the heterogeneous ASD compared with non-ASD and the most of the DMPs were notably hypermethylated. We observed these hypermethylated loci mapped to L1P, L1H, *AluJ*, and *AluS* elements, which are intermediate and youngest LINE-1, and oldest and intermediate *Alu*, respectively. This implies that hypermethylation suppressed the most active LINE-1 and *Alu* subfamilies (intermediate and young REs) in the heterogeneous ASD. However, the hypomethylation of intermediate and young of LINE-1 and *Alu* was shown in ASD with *CHD8* variants. Both hypermethylation and hypomethylation of REs interfere with gene expression of themselves and inserted genes in ASD (21, 25).

It is important to note that LINE-1 and *Alu* elements play important roles in human brain development and brain somatic mosaicism. LINE-1 and *Alu* elements can regulate nearby genes during brain development (52–55). LINE-1 and *Alu*

retrotransposition occurred more frequently in the brain than germline cells (56). Furthermore, Coufal's study, which compared LINE-1 activity in fetal neural progenitor cells (NPCs) to other somatic cells, revealed that NPCs have high retrotransposition of LINE-1 (57). They also discovered low DNAm at the LINE-1 promoter as well as a high copy number of LINE-1 in brain tissues when compared to other somatic cells (57, 58). Thus, we also performed an analysis in the validation cohort using data from post-mortem brain tissues (GSE131706), we identified significant 482 DMPs ($p < 0.05$) at LINE-1 and *Alu* elements. Interestingly, we found that all DMPs from the brain were overlapped with DMPs identified in the heterogenous ASD blood (see Additional file 10). These findings suggest that epigenetic dysregulation of LINE-1 and *Alu* elements in ASD may alter the function of autism-related genes regulated by these elements. To address this, we predicted the biological functions and networks of genes located nearby DMPs of LINE-1 and *Alu* elements. Neurological diseases and canonical pathways implicated in ASD were significantly associated with these genes (see Additional file 5). Moreover, interactome networks associated with ASD revealed several autism-related genes in the SFARI database (Figure 6). Especially, *Alu*Sg7 (cg16926147) within *KCNQ3* gene and L1PA3 (cg24094412) within *HCN1* gene that were hypermethylated and inversely related to aberrant gene expression in the blood and post-mortem brain tissues of several ASD cohort studies. Hypomethylated DMPs were also discovered in the most active *Alu* family, *AluY* (cg08998414), which is located on the *UBE2H* gene and has an inverse relationship with gene expression. *KCNQ3* encodes a protein involved in neuronal excitability; people with a *de novo* variant of this gene experience ASD features, and some were diagnosed with ASD (59). *HCN1* encodes a hyperpolarization-activated cation channel that is widely expressed in the brain regions (60). *HCN1* mutation causes epileptic encephalopathy and this mutation is associated with intellectual disability and autistic traits (61). *UBE2H* encodes an E2 ubiquitin-conjugating enzyme family protein that is involved in the protein ubiquitination mechanism. Genetic association and screening studies have shown that this gene is present in ASD individuals. (62, 63). Another interesting result is hypomethylation in the HAL1 family which was found exclusively in the heterogeneous ASD. HAL1 or half-L1 encodes only ORF1p which enhances the efficiency of their transposition, but the origin, biological properties, and subfamilies have not been well studied (64). HAL1 subfamilies were also not well annotated in our data. However, this result warrants further research of their biological activity in the ASD.

Here, we discovered LINE-1 and *Alu* methylation signatures in these genetically homogeneous ASD (both 16p11.2 del and *CHD8* variants). In the ASD with 16p11.2del, only locus-specific changes at LINE-1 and *Alu* elements were observed (Figure 4). We identified unique DMPs which target genes differentially expressed in the several ASD cohort studies including *AluY* within *XKR6* (cg21300361), *AluSq* within *ZNF107* (cg01772945), and L1MB3 within *MYEOV2* (cg13749477). These genes were genetic risk variants for ASD identified in genome-wide association study (GWAS), single nucleotide polymorphisms (SNPs), and copy number variation (CNV) studies (65–67). In the case of ASD with the *CHD8* variants, we observed a widespread reduction of LINE-1 and *Alu* methylation levels in global methylation and the active LINE-1 and *Alu* families (L1P, L1H, and *Alu*S). This change has far-reaching implications for even the oldest and fossil family (*AluJ* and FAM), as well as FRAM family. Furthermore, the unique LINE-1 and *Alu* methylation signatures of ASD with *CHD8* variants, such as L1MC5 (cg22706070) within *EHMT2*, *AluJo* (cg06421197) within *CASP1*, and *AluSx* (cg18699242, cg01963623, cg02169692) within *USP18*, were also hypomethylated. However, we found that these alterations are inconsistent with the heterogeneous ASD profile, in which most DMPs were hypermethylated. These results may be caused by disease-specific genetic variants of *CHD8* that is a huge difference from the ASD without any genetic variants or with undefined ones. CHD8 is a chromatin remodeling/modifier factor that plays a role in the transcription process required for brain development (10). LINE-1 and *Alu* elements have an activation and a repressive chromatin mark that is bound by several chromatin remodeling/modifier factors (52, 55, 68, 69). Aberrant *CHD8* function may be leading to changes in genome-wide epigenetic marks which can affect a variety of gene regulatory mechanisms. The inverse relationship between LINE-1/*Alu* methylation and gene expression were also observed in the ASD with *CHD8* variants. *EHMT2*, located nearby L1MC5, is a histone lysine methyltransferase involved with gene activation or repression. Gene and protein expression levels of *EHMT2* were significantly increased in the post-mortem brain tissues of ASD (70, 71). *CASP1* encodes cysteine-aspartic acid protease (caspase) enzyme involving apoptosis, monocyte cell fate, and immune response (72). *CASP1*, located nearby *AluJo*, was significantly elevated in the peripheral blood mononuclear cells of ASD (73), as well as overexpressed in two ASD studies including blood and post-mortem brain tissues. USP18 is a protein in the ubiquitin pathway which is essential for cell cycle, cell differentiation, and proliferation (74) and its CNV has been reported in ASD individuals (75). In transcriptome data obtained from several ASD studies, *USP18* was

significantly decreased, but not inversely related to hypomethylated positions of *AluSx* located in the upstream region of *USP18* gene. However, three probes (cg18699242, cg01963623, and cg27281093) at the same regions have been reported to be hypomethylated and they are the *CDH8* signature in the previous study (29). Our findings showed that DNAm of LINE-1 and *Alu* elements, located in the target genes, are connected with ASD-related genes. Moreover, biological functions and interactome of the genes located nearby LINE-1 and *Alu* methylation signatures in the ASD cohorts were associated with neurological diseases and developmental disorders, as well as canonical pathways implicated in ASD.

Unlike genetic changes, epigenetic alterations are not recorded in the genome and cannot be identified by genome sequencing. DNAm signatures are identified by comparing the methylation patterns of affected individuals to those typically developing control individuals. Several DNAm signatures have been established, and their effectiveness demonstrated as epigenetic markers for identifying variations of uncertain significance as pathogenic or benign (76). Although ASD pathogenesis occurs in the brain tissue, other systems such as the immune (77), metabolic (78), and gastrointestinal systems (79) are also affected in ASD individuals. DNAm in the blood is highly correlated to brain tissue samples and reflects environmental exposure (80). The discovery of distinct LINE-1 and *Alu* methylation signatures in ASD blood outlines their clinical potential to be used as non-invasive biomarkers. We conducted ROC curves analysis to predict a sensitivity and specificity of diagnosis with ASD using unique DMPs at LINE-1 and *Alu* elements identified in the blood of ASD individuals. Our findings show that LINE-1 and *Alu* methylation can be used to identify ASD with specific variants from unaffected individuals and classify them. However, additional research is required to determine its sensitivity and specificity in large and independent ASD cohorts.

In summary, we interrogated DNAm signatures of LINE-1 and *Alu* elements in the blood of ASD. DNAm signatures reflect environmental exposure linked to interaction with genetic factors through epigenetic processes. In the ASD cohort, DNAm signatures revealed changes in age acceleration residuals and LINE-1/*Alu* elements positions. The DNAm age was decelerated, meaning that an individual with ASD is epigenetically younger than their chronological age. Alteration of LINE-1 and *Alu* methylation induced a change in expression of genes located nearby these elements. Moreover, these genes have important functions related to neurodevelopment and cellular signaling. Disruption of these functions may lead to ASD features (Figure 10).

Limitations

Because of the limitation of available post-mortem brain tissues for each ASD variant in publicly available datasets, our analyses were carried out using methylation data from ASD blood. Further research with a large number of post-mortem brain tissues is required. We did not perform multiple test corrections in the analysis of the validation cohort using post-mortem brain tissues, due to the small sample size which affects the statistical analysis power. Moreover, methylome and transcriptome datasets used in our study were obtained from different cohorts. However, transcriptome data from several ASD studies may reflect the heterogeneity of ASD, and one of the ASD cohorts in our analyses is a heterogeneous group. It is important to note that changes in LINE-1 and *Alu* methylation may occur as a result of a genetic factor in the genetically homogeneous ASD.

Conclusions

Locus-specific DNAm of LINE-1 and *Alu* elements in ASD, as well as its functional impact on gene regulation, were firstly reported in our study. Our analyses revealed LINE-1 and *Alu* methylation changes in a locus- and family-specific manner which were different according to each ASD cohort. By integrating methylome and transcriptome data, the target genes of LINE-1 and *Alu* elements were identified. These target genes were differentially expressed in multiple ASD cohorts, and their functions were related to neurological diseases and developmental disorders such as ASD. Our research also demonstrated that the LINE-1 and *Alu* signatures can be applied to diagnose and classify people with ASD. Finally, our finding will provide a better understanding of the impact of LINE-1 and *Alu* elements in ASD, at least in the blood. Our study provides evidence supporting future studies on the role of LINE-1 and *Alu* related to ASD neuropathology using human post-mortem brain tissues. However,

further functional studies will be necessary to investigate the subsequent impact upon the target genes and fully elucidate the role of REs in ASD biology.

Abbreviations

16p11.2del
16p11.2 deletion
3'UTR
3' untranslated region
5'UTR
5' untranslated region
ASD
Autism spectrum disorder
CASP1
Caspase 1
CHD8
Chromodomain helicase DNA-binding protein 8
CNV
Copy number variation
DEGs
Differentially expressed genes
DMGs
Differentially methylated genes
DMPs
Differentially methylated positions
DMRs
Differentially methylated regions
DNAm
DNA methylation
EAA
Epigenetic age acceleration
EHMT2
Euchromatic Histone Lysine Methyltransferase 2
FDR
False discovery rate
FNBP1
Formin binding protein 1
GWAS
Genome-wide association study
HCN1
Hyperpolarization activated cyclic nucleotide gated potassium channel 1
KCNQ3
Potassium voltage-gated channel subfamily Q member 3
LINE-1
Long interspersed nucleotide element-1
MeCp2
Methyl-CpG binding protein 2
MYEOV2

Myeloma overexpressed 2
NDST1
N-deacetylase and N-sulfotransferase 1
ROC
Receiver operating characteristic
SNPs
Single nucleotide polymorphisms
TSS
Transcriptional start site
UBE2H
Ubiquitin Conjugating Enzyme E2 H
USP18
Ubiquitin-specific peptidase 18
USP6
Ubiquitin specific peptidase 6
XKR6
XK related 6
ZNF107
Zinc Finger Protein 107

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests

Funding

This study was supported by Ratchadapisek Somphot Fund for Supporting Research Unit, Chulalongkorn University (grant number GRU 6300437001-1 and GRU_64_033_37_004) and the National Research University Project, Office of Higher Education Commission (NRU59-031-HR), awarded to Tewarit Sarachana), Postdoctoral Fellowship Program, Siriraj Hospital, Mahidol University (awarded to Chanachai Sae-Lee), and the Second Century Fund (C2F), Chulalongkorn University (awarded to Thanit Saeliw).

Authors' contributions

T.Sae. performed transcriptome analysis, biological function, and interactome prediction, and drafted the manuscript under the supervision of T.S., T.T. and C.S. C.S., N.I., D.G., and T.P. performed DNAm analysis for LINE-1 and *Alu* elements. T.S.

supervised transcriptome analysis and interpretation. T.S. supervised the biological function and interactome prediction using IPA. C.S. had primary responsibility for final content. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank Dr. Hyang-Min Byun and Dr. Timothy M. Barrow for their assistance with the analysis of DNA methylation of REs.

References

1. Association AP. Diagnostic and Statistical Manual of Mental Disorders. 5th edition. 2013.
2. Maenner MJ, Shaw KA, Baio J, EdS, Washington A, Patrick M, et al. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. MMWR Surveill Summ. 2020;69(4):1-12.
3. Schaefer GB, Mendelsohn NJ. Genetics evaluation for the etiologic diagnosis of autism spectrum disorders. Genet Med. 2008;10(1):4-12.
4. Bolte S, Girdler S, Marschik PB. The contribution of environmental exposure to the etiology of autism spectrum disorder. Cell Mol Life Sci. 2019;76(7):1275-97.
5. Stamou M, Streifel KM, Goines PE, Lein PJ. Neuronal connectivity as a convergent target of gene x environment interactions that confer risk for Autism Spectrum Disorders. Neurotoxicol Teratol. 2013;36:3-16.
6. Tordjman S, Somogyi E, Coulon N, Kermarrec S, Cohen D, Bronsard G, et al. Gene x Environment interactions in autism spectrum disorders: role of epigenetic mechanisms. Front Psychiatry. 2014;5:53.
7. Folstein SE, Rosen-Sheidley B. Genetics of autism: complex aetiology for a heterogeneous disorder. Nat Rev Genet. 2001;2(12):943-55.
8. Merner N, Forgeot d'Arc B, Bell SC, Maussion G, Peng H, Gauthier J, et al. A de novo frameshift mutation in chromodomain helicase DNA-binding domain 8 (CHD8): A case report and literature review. Am J Med Genet A. 2016;170A(5):1225-35.
9. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, et al. Association between microdeletion and microduplication at 16p11.2 and autism. N Engl J Med. 2008;358(7):667-75.
10. Weissberg O, Elliott E. The Mechanisms of CHD8 in Neurodevelopment and Autism Spectrum Disorders. Genes (Basel). 2021;12(8).
11. Handy DE, Castro R, Loscalzo J. Epigenetic modifications: basic mechanisms and role in cardiovascular disease. Circulation. 2011;123(19):2145-56.
12. Zaidi SK, Young DW, Montecino M, Lian JB, Stein JL, van Wijnen AJ, et al. Architectural epigenetics: mitotic retention of mammalian transcriptional regulatory information. Mol Cell Biol. 2010;30(20):4758-66.
13. Edgar RD, Jones MJ, Meaney MJ, Turecki G, Kobor MS. BECon: a tool for interpreting DNA methylation findings from blood in the context of brain. Transl Psychiatry. 2017;7(8):e1187.
14. van de Haar HJ, Burgmans S, Jansen JF, van Osch MJ, van Buchem MA, Muller M, et al. Blood-Brain Barrier Leakage in Patients with Early Alzheimer Disease. Radiology. 2016;281(2):527-35.
15. Andrews SV, Sheppard B, Windham GC, Schieve LA, Schendel DE, Croen LA, et al. Case-control meta-analysis of blood DNA methylation and autism spectrum disorder. Mol Autism. 2018;9:40.
16. Garcia-Perez JL, Widmann TJ, Adams IR. The impact of transposable elements on mammalian development. Development. 2016;143(22):4101-14.
17. Britten RJ, Baron WF, Stout DB, Davidson EH. Sources and evolution of human Alu repeated sequences. Proc Natl Acad Sci U S A. 1988;85(13):4770-4.

18. Smit AF, Toth G, Riggs AD, Jurka J. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. J Mol Biol. 1995;246(3):401-17.
19. Khan H, Smit A, Boissinot S. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. Genome Res. 2006;16(1):78-87.
20. Mills RE, Bennett EA, Iskow RC, Devine SE. Which transposable elements are active in the human genome? Trends Genet. 2007;23(4):183-91.
21. Saeliw T, Tangsuwansri C, Thongkorn S, Chonchaiya W, Suphapeetiporn K, Mutirangura A, et al. Integrated genome-wide Alu methylation and transcriptome profiling analyses reveal novel epigenetic regulatory networks associated with autism spectrum disorder. Mol Autism. 2018;9:27.
22. Sae-Lee C, Biasi J, Robinson N, Barrow TM, Mathers JC, Koutsidis G, et al. DNA methylation patterns of LINE-1 and Alu for pre-symptomatic dementia in type 2 diabetes. PLoS One. 2020;15(6):e0234578.
23. Barrow TM, Wong Doo N, Milne RL, Giles GG, Willmore E, Strathdee G, et al. Analysis of retrotransposon subfamily DNA methylation reveals novel early epigenetic changes in chronic lymphocytic leukemia. Haematologica. 2021;106(1):98-110.
24. Jacob-Hirsch J, Eyal E, Knisbacher BA, Roth J, Cesarkas K, Dor C, et al. Whole-genome sequencing reveals principles of brain retrotransposition in neurodevelopmental disorders. Cell Res. 2018;28(2):187-203.
25. Shpyleva S, Melnyk S, Pavliv O, Pogribny I, Jill James S. Overexpression of LINE-1 Retrotransposons in Autism Brain. Mol Neurobiol. 2018;55(2):1740-9.
26. Tangsuwansri C, Saeliw T, Thongkorn S, Chonchaiya W, Suphapeetiporn K, Mutirangura A, et al. Investigation of epigenetic regulatory networks associated with autism spectrum disorder (ASD) by integrated global LINE-1 methylation and gene expression profiling analyses. PLoS One. 2018;13(7):e0201071.
27. Garcia-Ortiz MV, de la Torre-Aguilar MJ, Morales-Ruiz T, Gomez-Fernandez A, Flores-Rojas K, Gil-Campos M, et al. Analysis of Global and Local DNA Methylation Patterns in Blood Samples of Patients With Autism Spectrum Disorder. Front Pediatr. 2021;9:685310.
28. Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, et al. NCBI GEO: archive for functional genomics data sets--10 years on. Nucleic Acids Res. 2011;39(Database issue):D1005-10.
29. Siu MT, Butcher DT, Turinsky AL, Cytrynbaum C, Stavropoulos DJ, Walker S, et al. Functional DNA methylation signatures for autism spectrum disorder genomic risk loci: 16p11.2 deletions and CHD8 variants. Clin Epigenetics. 2019;11(1):103.
30. Corley MJ, Vargas-Maya N, Pang APS, Lum-Jones A, Li D, Khadka V, et al. Epigenetic Delay in the Neurodevelopmental Trajectory of DNA Methylation States in Autism Spectrum Disorders. Front Genet. 2019;10:907.
31. Pelegi-Siso D, de Prado P, Ronkainen J, Bustamante M, Gonzalez JR. methylclock: a Bioconductor package to estimate DNA methylation age. Bioinformatics. 2021;37(12):1759-60.
32. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013;14(10):R115.
33. https: //www.who.int/maternal_child_adolescent/adolescence/en/. 2019.
34. Fortin JP, Triche TJ, Jr., Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Bioinformatics. 2017;33(4):558-60.
35. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. Behav Brain Res. 2001;125(1-2):279-84.
36. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, et al. TM4 microarray software suite. Methods Enzymol. 2006;411:134-93.
37. Jalili V, Afgan E, Gu Q, Clements D, Blankenberg D, Goecks J, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. Nucleic Acids Res. 2020;48(W1):W395-W402.
38. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884-i90.

39. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12(4):357-60.
40. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923-30.
41. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
42. Kramer A, Green J, Pollard J, Jr., Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics. 2014;30(4):523-30.
43. Srivastava S, Sahin M. Autism spectrum disorder and epileptic encephalopathy: common causes, many questions. J Neurodev Disord. 2017;9:23.
44. Bell L, Wittkowski A, Hare DJ. Movement Disorders and Syndromic Autism: A Systematic Review. J Autism Dev Disord. 2019;49(1):54-67.
45. Bell CG, Lowe R, Adams PD, Baccarelli AA, Beck S, Bell JT, et al. DNA methylation aging clocks: challenges and recommendations. Genome Biol. 2019;20(1):249.
46. Wu X, Chen W, Lin F, Huang Q, Zhong J, Gao H, et al. DNA methylation profile is a quantitative measure of biological aging in children. Aging (Albany NY). 2019;11(22):10031-51.
47. Hu VW, Sarachana T, Kim KS, Nguyen A, Kulkarni S, Steinberg ME, et al. Gene expression profiling differentiates autism case-controls and phenotypic variants of autism spectrum disorders: evidence for circadian rhythm dysfunction in severe autism. Autism Res. 2009;2(2):78-97.
48. Hu VW, Steinberg ME. Novel clustering of items from the Autism Diagnostic Interview-Revised to define phenotypes within autism spectrum disorders. Autism Res. 2009;2(2):67-77.
49. Lee EC, Hu VW. Phenotypic Subtyping and Re-Analysis of Existing Methylation Data from Autistic Probands in Simplex Families Reveal ASD Subtype-Associated Differentially Methylated Genes and Biological Functions. Int J Mol Sci. 2020;21(18).
50. Frappart PO, McKinnon PJ. Mouse models of DNA double-strand break repair and neurological disease. DNA Repair (Amst). 2008;7(7):1051-60.
51. Pornthanakasem W, Kongruttanachok N, Phuangphairoj C, Suyarnsestakorn C, Sanghangthum T, Oonsiri S, et al. LINE-1 methylation status of endogenous DNA double-strand breaks. Nucleic Acids Res. 2008;36(11):3667-75.
52. Jonsson ME, Ludvik Brattas P, Gustafsson C, Petri R, Yudovich D, Pircs K, et al. Activation of neuronal genes via LINE-1 elements upon global DNA demethylation in human neural progenitors. Nat Commun. 2019;10(1):3182.
53. Polak P, Domany E. Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. BMC Genomics. 2006;7:133.
54. Batzer MA, Deininger PL. Alu repeats and human genomic diversity. Nat Rev Genet. 2002;3(5):370-9.
55. Su M, Han D, Boyd-Kirkup J, Yu X, Han JJ. Evolution of Alu elements toward enhancers. Cell Rep. 2014;7(2):376-85.
56. Upton KR, Baillie JK, Faulkner GJ. Is somatic retrotransposition a parasitic or symbiotic phenomenon? Mob Genet Elements. 2011;1(4):279-82.
57. Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, et al. L1 retrotransposition in human neural progenitor cells. Nature. 2009;460(7259):1127-31.
58. Muotri AR. L1 Retrotransposition in Neural Progenitor Cells. Methods Mol Biol. 2016;1400:157-63.
59. Sands TT, Miceli F, Lesca G, Beck AE, Sadleir LG, Arrington DK, et al. Autism and developmental disability caused by KCNQ3 gain-of-function variants. Ann Neurol. 2019;86(2):181-92.
60. Chang X, Wang J, Jiang H, Shi L, Xie J. Hyperpolarization-Activated Cyclic Nucleotide-Gated Channels: An Emerging Role in Neurodegenerative Diseases. Front Mol Neurosci. 2019;12:141.

61. Nava C, Dalle C, Rastetter A, Striano P, de Kovel CG, Nabbout R, et al. De novo mutations in HCN1 cause early infantile epileptic encephalopathy. Nat Genet. 2014;46(6):640-5.
62. Martin I, Vourc'h P, Mahe M, Thepault RA, Antar C, Vedrine S, et al. Association study of the ubiquitin conjugating enzyme gene UBE2H in sporadic ALS. Amyotroph Lateral Scler. 2009;10(5-6):432-5.
63. Autism Spectrum Disorders Working Group of The Psychiatric Genomics C. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. Mol Autism. 2017;8:21.
64. Bao W, Jurka J. Origin and evolution of LINE-1 derived "half-L1" retrotransposons (HAL1). Gene. 2010;465(1-2):9-16.
65. Thomas TR, Koomar T, Casten L, Tener AJ, Bahl E, Michaelson JJ. Clinical autism subscales have common genetic liability that is heritable, pleiotropic, and generalizable to the general population. medRxiv. 2021.
66. Yingjun X, Haiming Y, Mingbang W, Liangying Z, Jiaxiu Z, Bing S, et al. Copy number variations independently induce autism spectrum disorder. Biosci Rep. 2017;37(4).
67. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;51(3):431-44.
68. Ferrari R, de Llobet Cucalon LI, Di Vona C, Le Dilly F, Vidal E, Lioutas A, et al. TFIIIC Binding to Alu Elements Controls Gene Expression via Chromatin Looping and Histone Acetylation. Mol Cell. 2020;77(3):475-87 e11.
69. He J, Fu X, Zhang M, He F, Li W, Abdul MM, et al. Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells. Nat Commun. 2019;10(1):34.
70. Wang ZJ, Zhong P, Ma K, Seo JS, Yang F, Hu Z, et al. Amelioration of autism-like social deficits by targeting histone methyltransferases EHMT1/2 in Shank3-deficient mice. Mol Psychiatry. 2020;25(10):2517-33.
71. Balan S, Iwayama Y, Maekawa M, Toyota T, Ohnishi T, Toyoshima M, et al. Exon resequencing of H3K9 methyltransferase complex genes, EHMT1, EHTM2 and WIZ, in Japanese autism subjects. Mol Autism. 2014;5(1):49.
72. Denes A, Lopez-Castejon G, Brough D. Caspase-1: is IL-1 just the tip of the ICEberg? Cell Death Dis. 2012;3:e338.
73. Siniscalco D, Sapone A, Giordano C, Cirillo A, de Novellis V, de Magistris L, et al. The expression of caspases is enhanced in peripheral blood mononuclear cells of autism spectrum disorder patients. J Autism Dev Disord. 2012;42(7):1403-10.
74. Honke N, Shaabani N, Zhang DE, Hardt C, Lang KS. Multiple functions of USP18. Cell Death Dis. 2016;7(11):e2444.
75. Chen CH, Chen HI, Chien WH, Li LH, Wu YY, Chiu YN, et al. High resolution analysis of rare copy number variants in patients with autism spectrum disorder from Taiwan. Sci Rep. 2017;7(1):11919.
76. Chater-Diehl E, Goodman SJ, Cytrynbaum C, Turinsky AL, Choufani S, Weksberg R. Anatomy of DNA methylation signatures: Emerging insights and applications. Am J Hum Genet. 2021;108(8):1359-66.
77. Meltzer A, Van de Water J. The Role of the Immune System in Autism Spectrum Disorder. Neuropsychopharmacology. 2017;42(1):284-98.
78. Cheng N, Rho JM, Masino SA. Metabolic Dysfunction Underlying Autism Spectrum Disorder and Potential Treatment Approaches. Front Mol Neurosci. 2017;10:34.
79. Wasilewska J, Klukowski M. Gastrointestinal symptoms and autism spectrum disorder: links and risks - a possible new overlap syndrome. Pediatric Health Med Ther. 2015;6:153-66.
80. Tylee DS, Kawaguchi DM, Glatt SJ. On the outside, looking in: a review and evaluation of the comparability of blood and brain "-omes". Am J Med Genet B Neuropsychiatr Genet. 2013;162B(7):595-603.

Figures

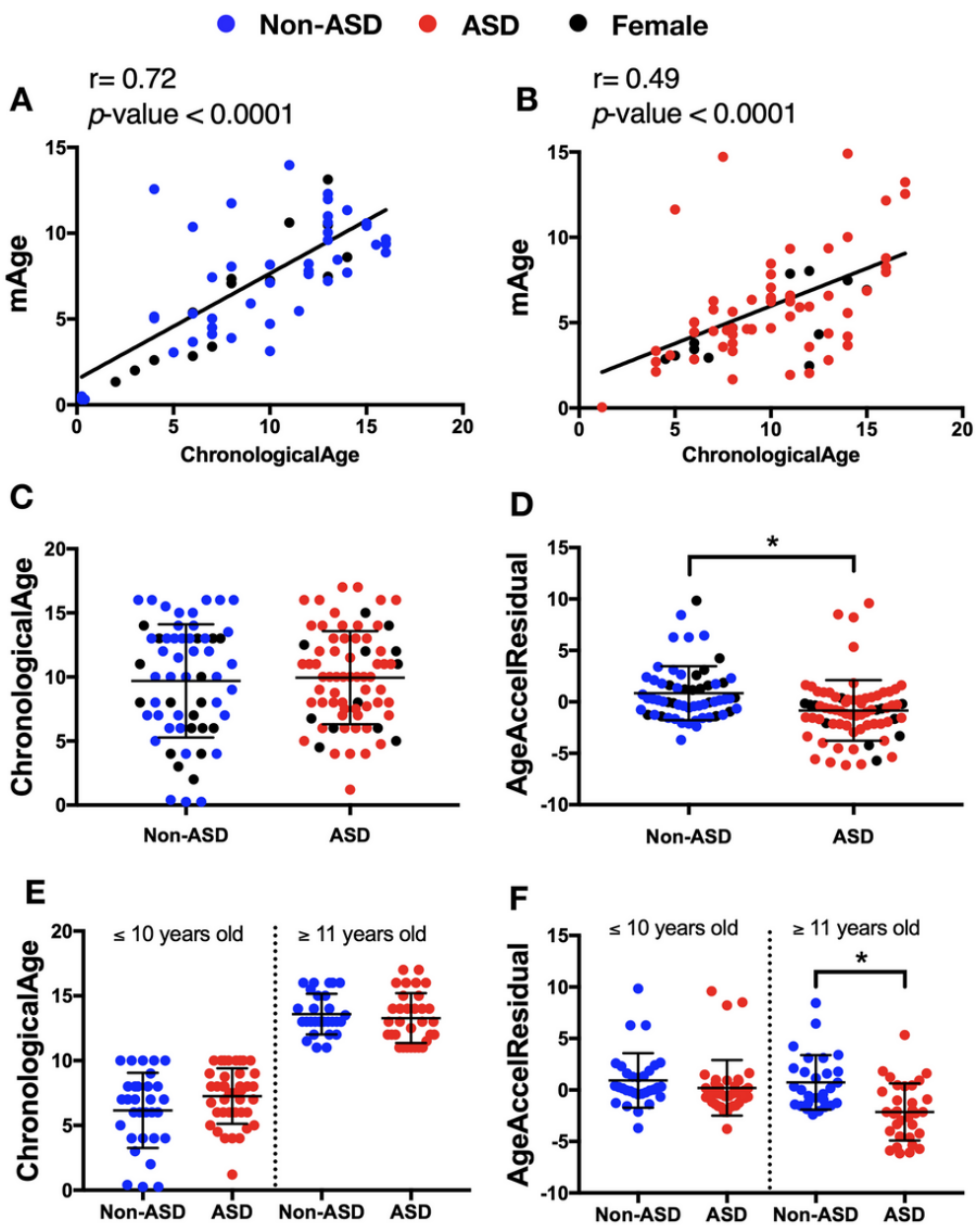


Figure 1

Age acceleration residual changes in ASD. Correlation of DNAm Age and chronological age in non-ASD (n = 59) (A) and ASD (n = 72) (B). Differences of the chronological age (C) and age acceleration residual (AgeAccelResidual) (D) between non-ASD and ASD. Differences of the chronological age (E) and AgeAccelResidual (F) between non-ASD and ASD by ≤ 10 years old and ≥ 11 years old. Mean ± SD. *$p < 0.05$.

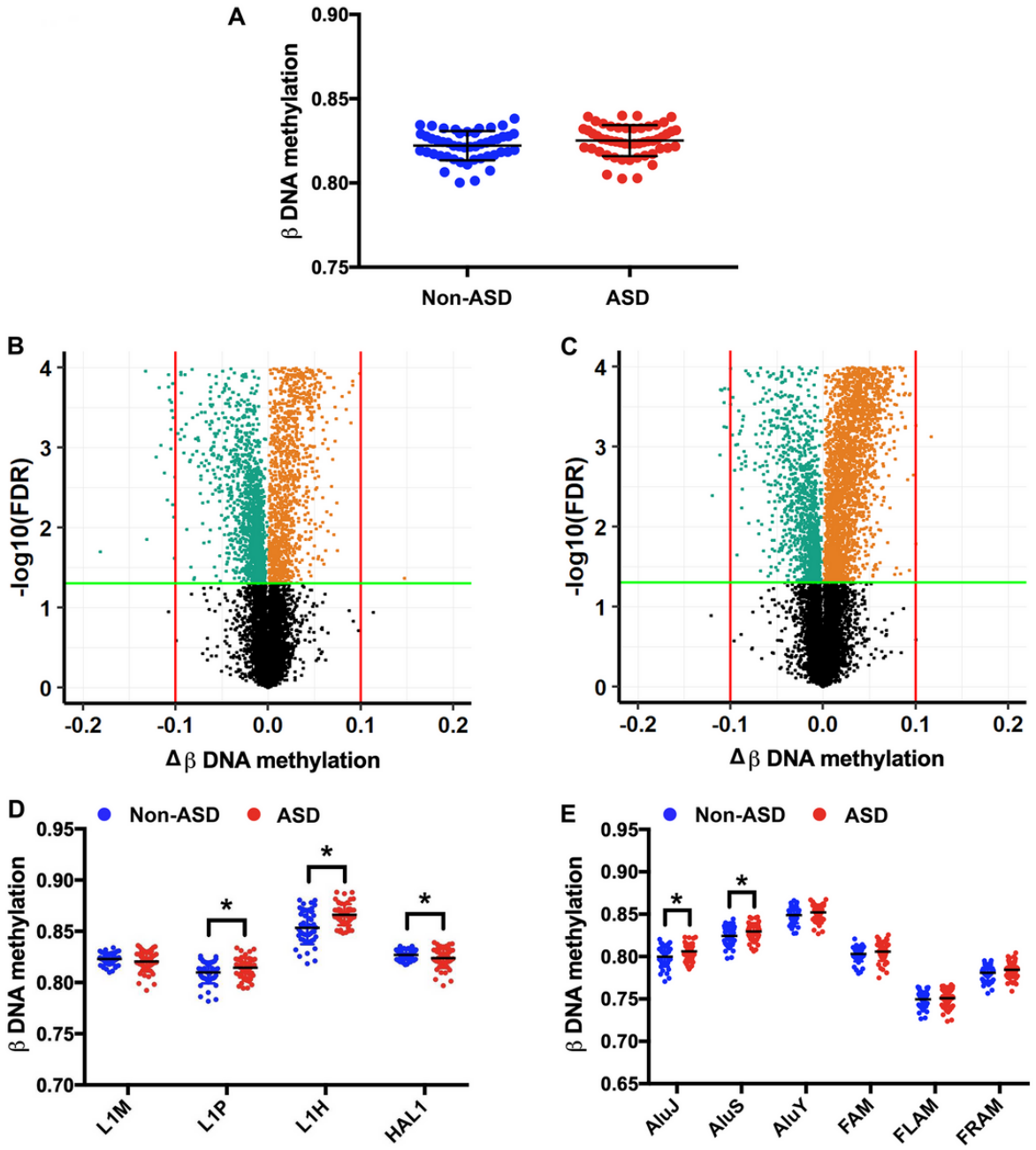


Figure 2

Methylation of repetitive elements (LINE-1 and *Alu*) in non-ASD (n = 48) and ASD (n = 52). Global DNA methylation (A). Volcano plots of mean change in methylation (Δβ) of LINE-1 (B) and *Alu* (C) against -log10 FDR-adjusted p-value (P_{FDR}) of ASD compared with non-ASD; the green line represents P_{FDR} = 0.05, the red line represents 10% of methylation changes, green dots represent hypomethylation loci, and orange dots represent hypermethylation loci. Changes in DNA methylation (Δβ) of ASD compared with non-ASD by subfamily of LINE-1 (D) and *Alu* (E). Mean ± SD. *p < 0.05.

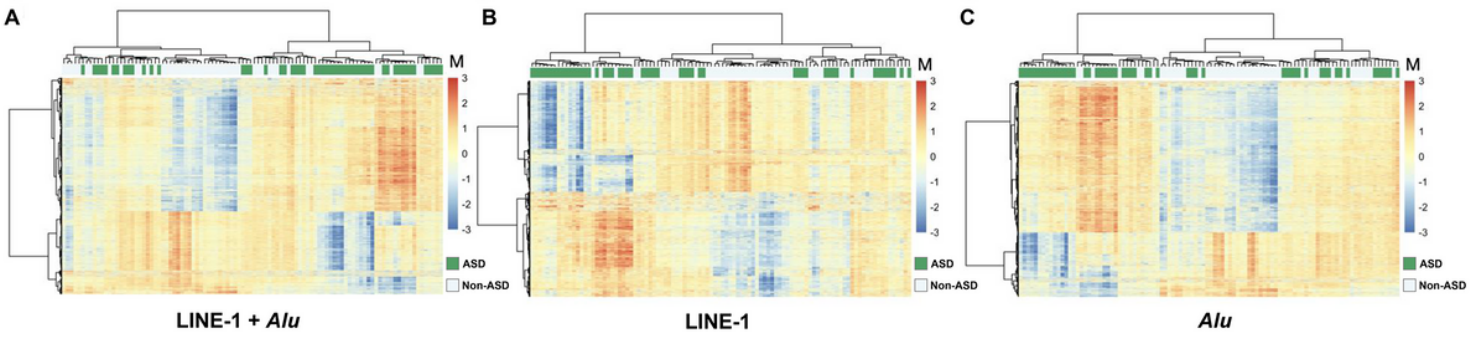


Figure 3

Unsupervised hierarchical cluster heatmap of the significant differentially methylated loci of the repetitive elements in non-ASD and ASD. Clustering of the 7,165 significant differentially methylated loci (combining of LINE-1 and *Alu* loci) (A). Clustering of the 2,802 significant differentially methylated loci (LINE-1) (B). Clustering of the 4,363 significant differentially methylated loci (*Alu*) (C). The color scale indicates methylation level (M value), from low (blue) to high (red). Green color represents ASD and white color represents non-ASD.

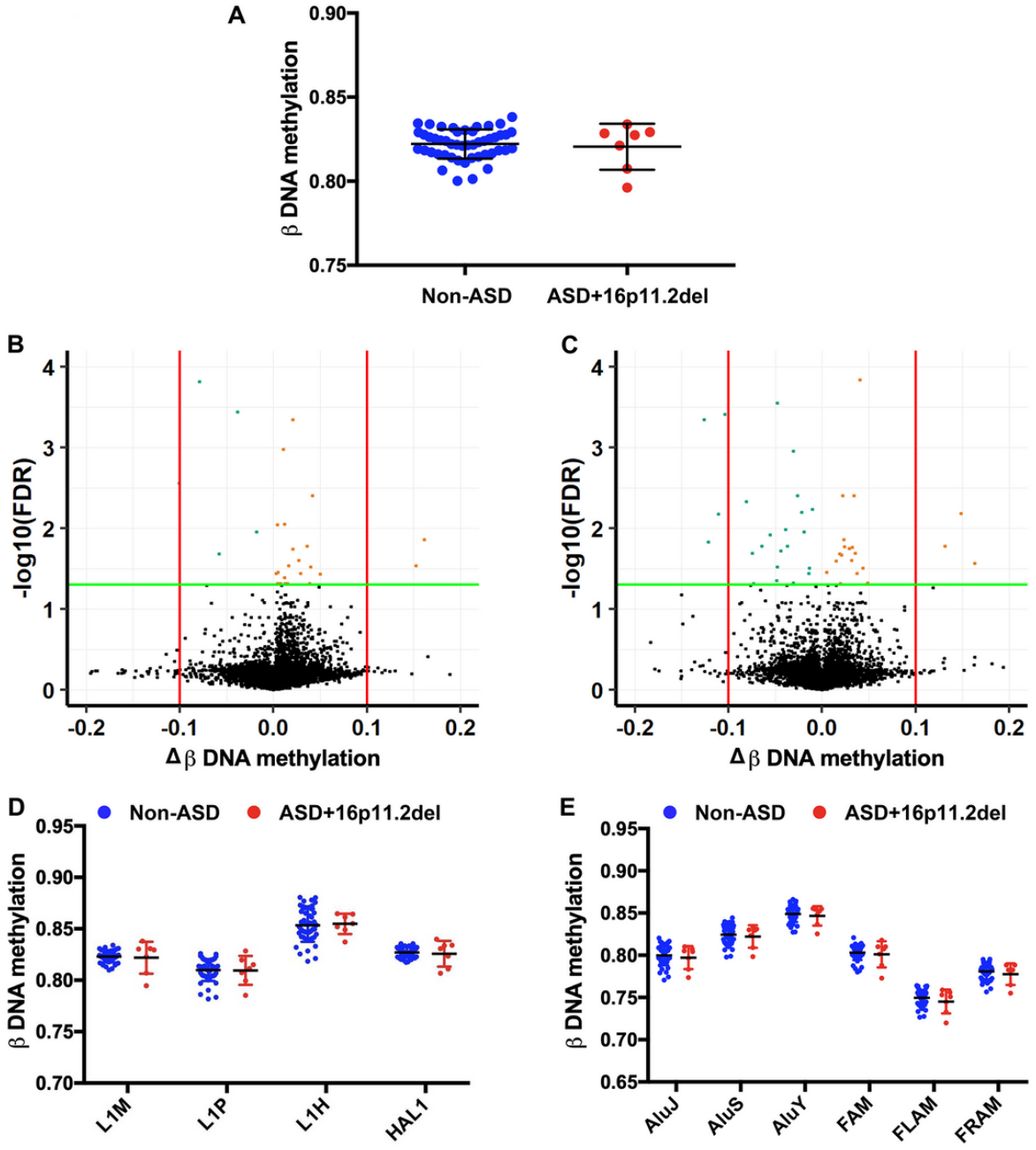


Figure 4

Methylation of repetitive elements (LINE-1 and *Alu*) in non-ASD (n = 48) and ASD patients who carry 16p11.2 deletions (n = 7). Global DNA methylation (A). Volcano plots of mean change in methylation (Δβ) of LINE-1 (B) and *Alu* (C) against -log10 FDR-adjusted p-value (P_{FDR}) of ASD compared with non-ASD; the green line represents P_{FDR} = 0.05, the red line represents 10% of methylation changes, green dots represent hypomethylation loci, and orange dots represent hypermethylation loci. Changes in DNA methylation (Δβ) of ASD with 16p11.2 deletion compared with non-ASD by subfamily of LINE-1 (D) and *Alu* (E). Mean ± SD.

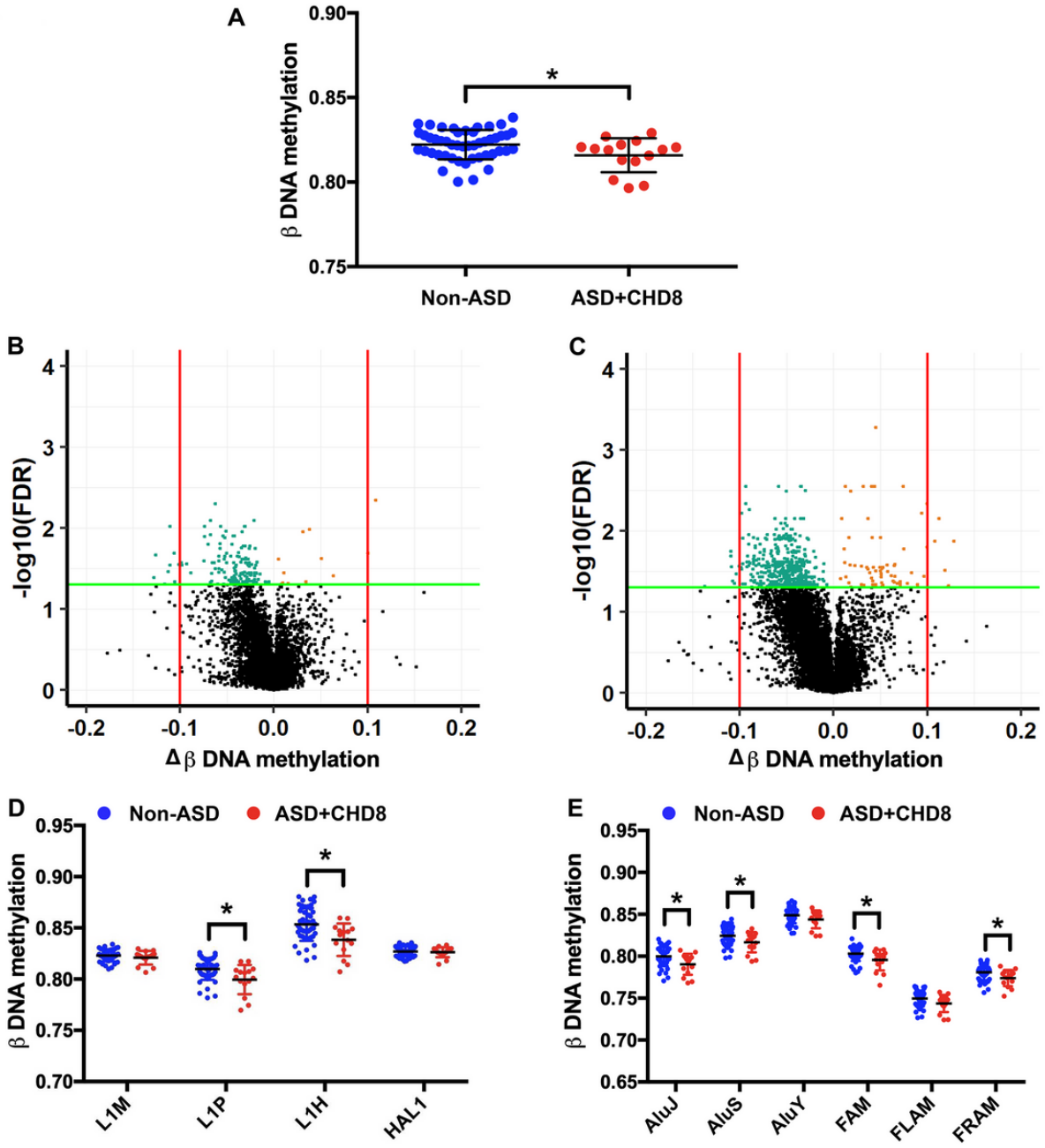
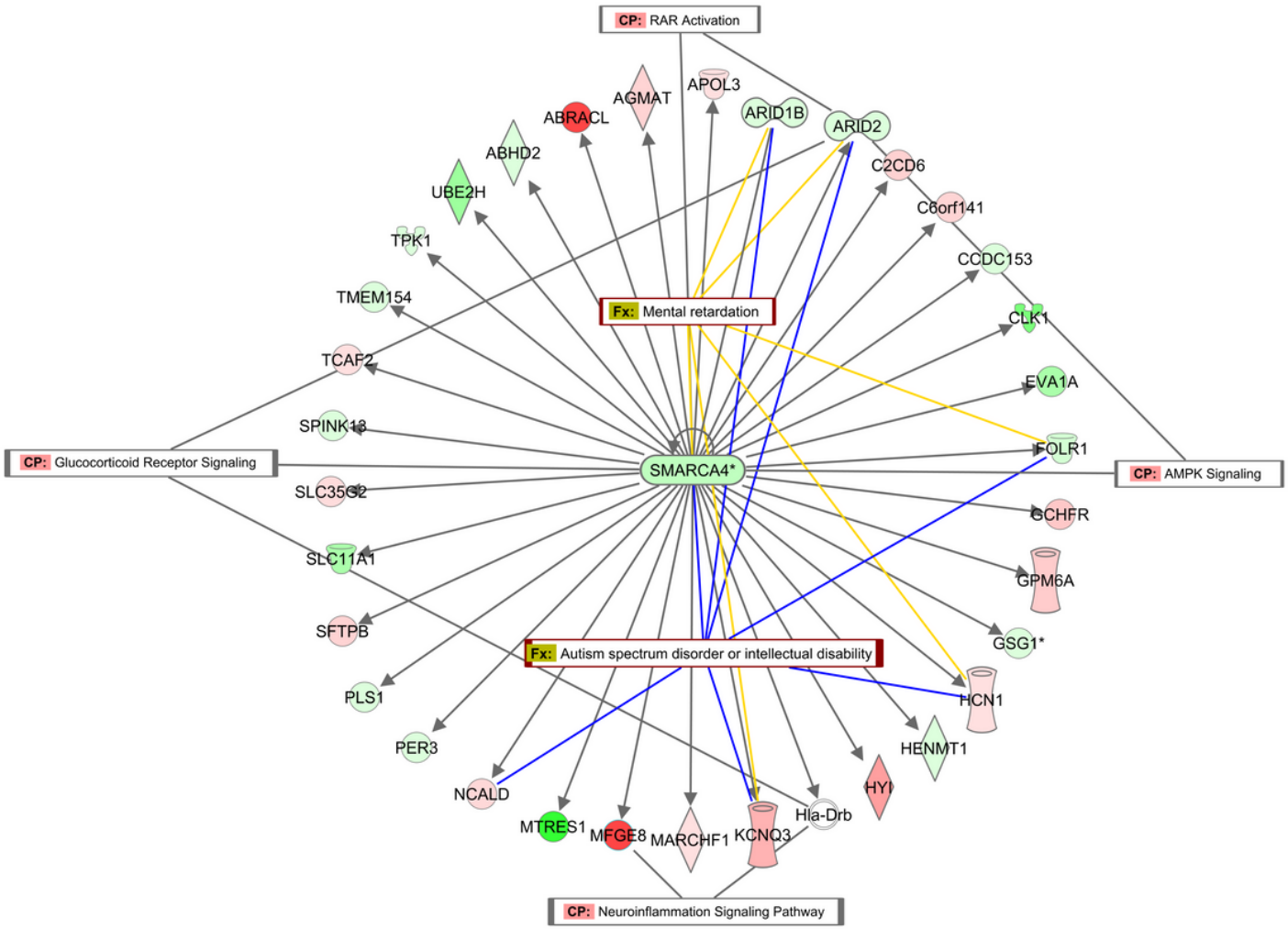


Figure 5

Methylation of repetitive elements (LINE-1 and *Alu*) in non-ASD (n = 48) and ASD patients who carry *CHD8* variants (n = 15). Global DNA methylation (A). Volcano plots of mean change in methylation (Δβ) of LINE-1 (B) and *Alu* (C) against -log10 FDR-adjusted p-value (P_{FDR}) of ASD compared with non-ASD; the green line represents P_{FDR} = 0.05, the red line represents 10% of methylation changes, green dots represent hypomethylation loci, and orange dots represent hypermethylation loci. Changes in DNA methylation (Δβ) of ASD with *CHD8* compared with non-ASD by subfamily of LINE-1 (D) and *Alu* (E). Mean ± SD. *p < 0.05.



© 2000-2021 QIAGEN. All rights reserved.

Figure 6

The regulatory network of differentially methylated genes (DMGs) in heterogeneous ASD that is related to neurological diseases. The gene regulatory network was predicted by ingenuity pathway analysis software using the list of DMGs (colored; red = hypermethylation; green = hypomethylation)

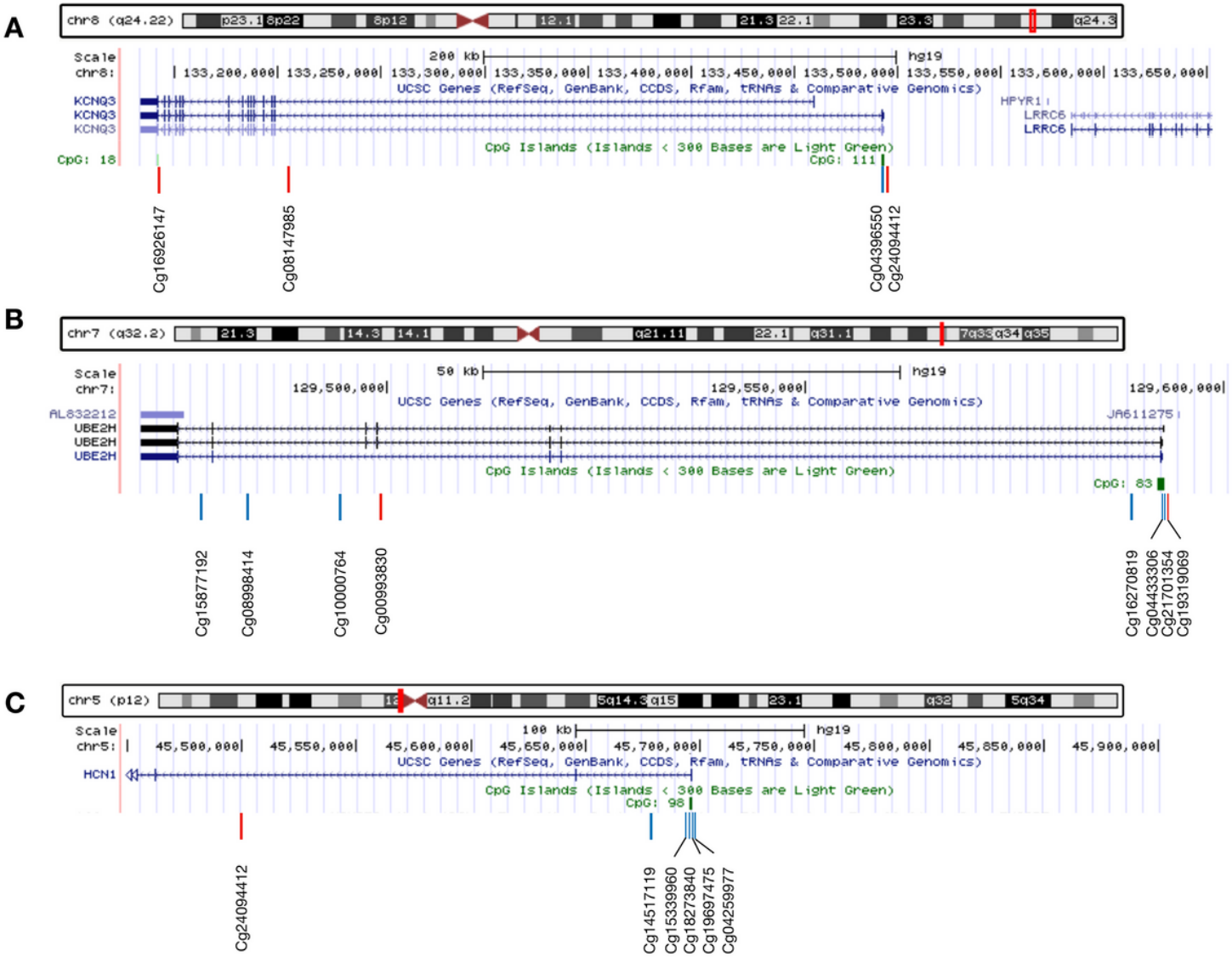


Figure 7

Genomic location of significant differentially methylated retrotransposons in heterogeneous ASD. Genomic location on *potassium voltage-gated channel subfamily Q member 3* (*KCNQ3*) (A), *ubiquitin conjugating enzyme E2 H* (*UBE2H*) (B), and *hyperpolarization activated cyclic nucleotide gated potassium channel 1* (*HCN1*) (C): blue line represents hypomethylation, red line represents hypermethylation.

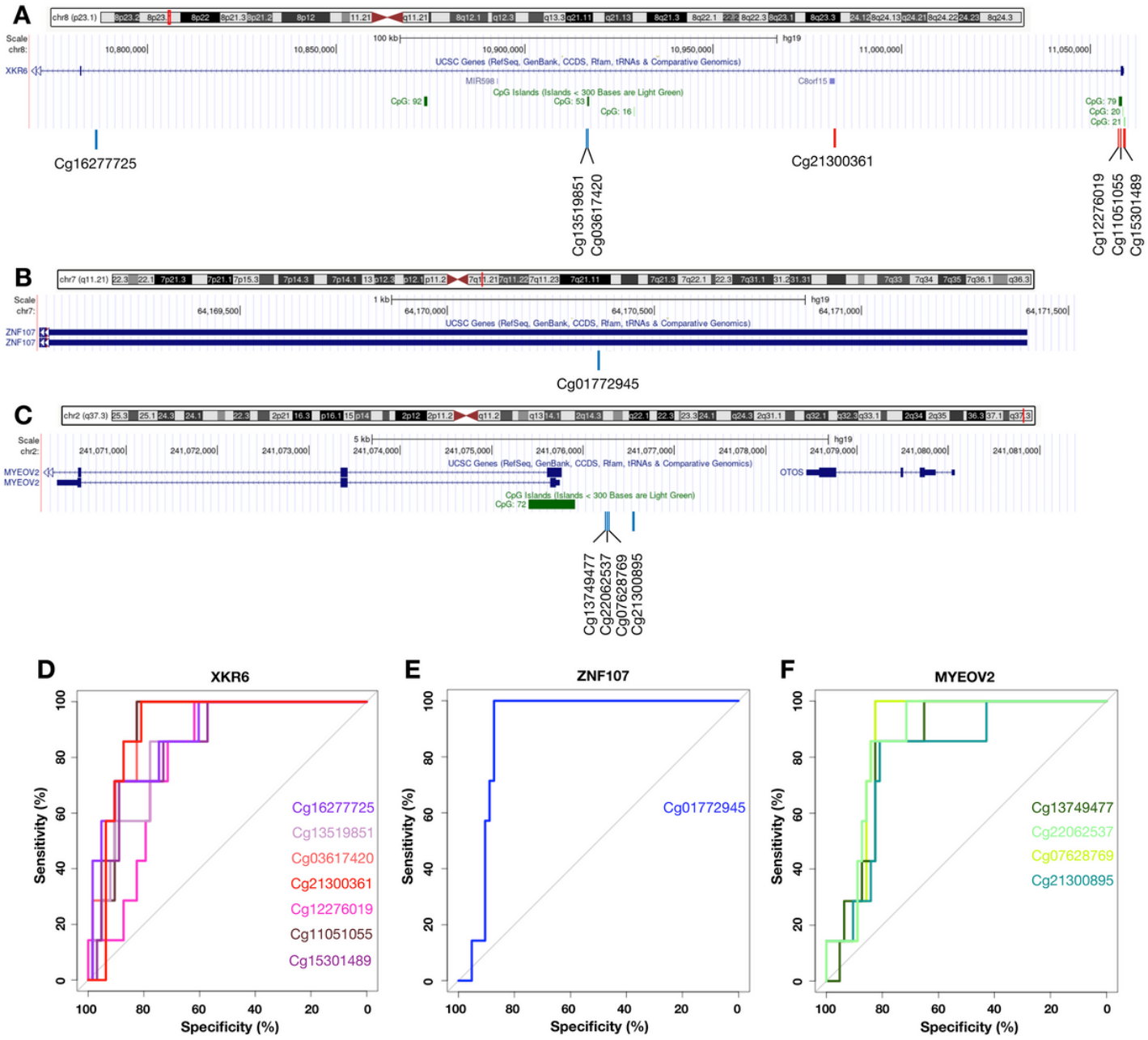


Figure 8

Genomic location and specificity of the unique differentially methylated regions in ASD with 16p11.2 deletion. Genomic location on *XK related 6 (XKR6)* (A), *Zinc Finger Protein 107 (ZNF107)* (B) and *Myeloma overexpressed 2 (MYEOV2)* (C): blue line represents hypomethylation, red line represents hypermethylation. The receiver operating characteristic (ROC) analysis of the unique differentially methylated regions of ASD with 16p11.2 deletion (n = 7) was performed against non-ASD (n = 48) and ASD with *CHD8* variants (n = 15). Specificity and sensitivity of the unique differentially methylated regions for *XKR6* (D), *ZNF107* (E) and *MYEOV2* (F).

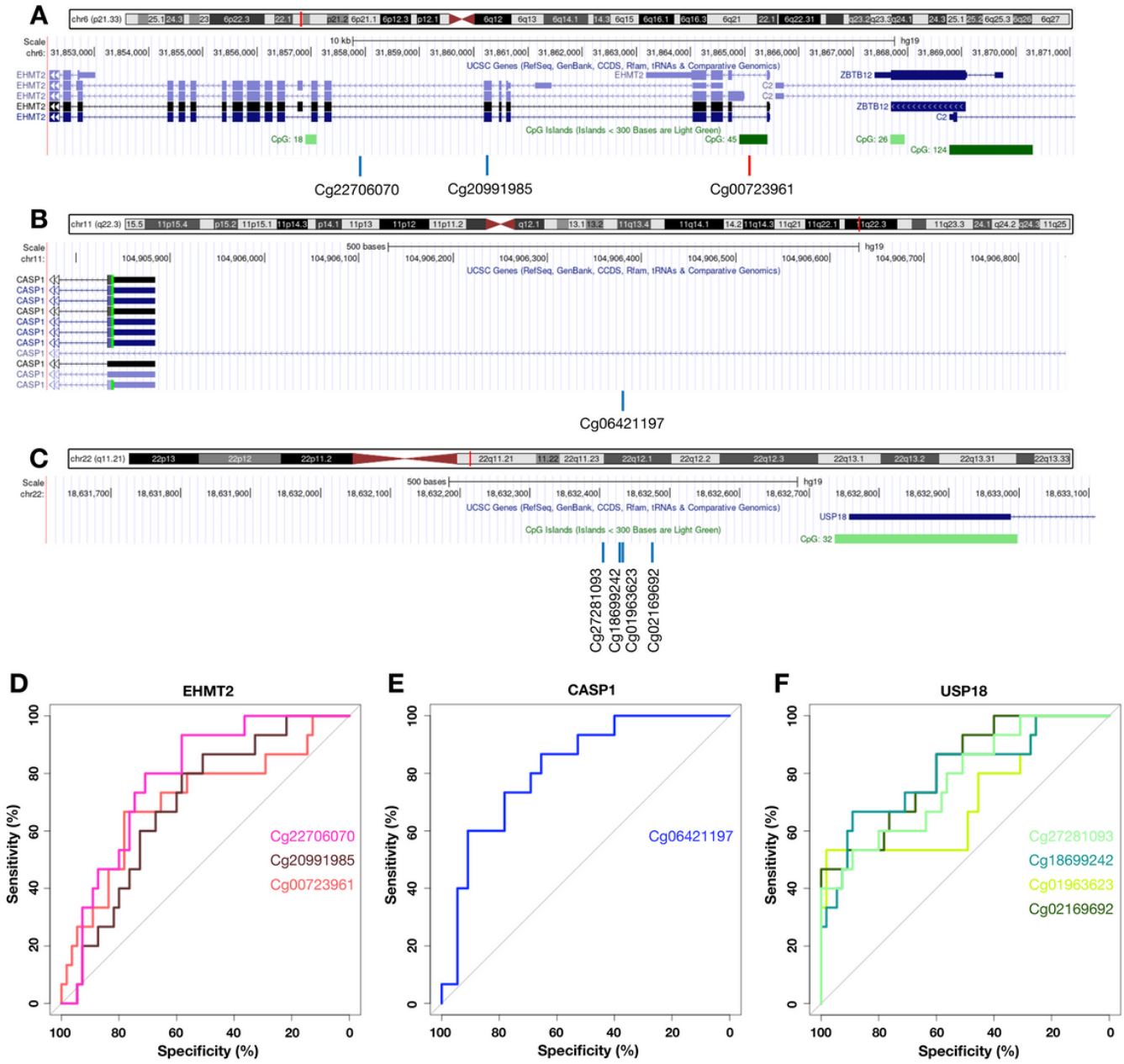


Figure 9

Genomic location and specificity of the unique differentially methylated regions in ASD with *CHD8* variants. Genomic location on *Euchromatic Histone Lysine Methyltransferase 2 (EHMT2)* (A), *Caspase 1 (CASP1)* (B) and *Ubiquitin Specific Peptidase 18 (USP18)* (C): blue line represents hypomethylation, red line represents hypermethylation. The receiver operating characteristic (ROC) analysis of the unique differentially methylated regions of ASD with *CHD8* variants (n = 15) was performed against non-ASD (n = 48) and ASD with 16p11.2 deletion (n = 7). Specificity and sensitivity of the unique differentially methylated regions for *EHMT2* (D), *CASP1* (E) and *USP18* (F).

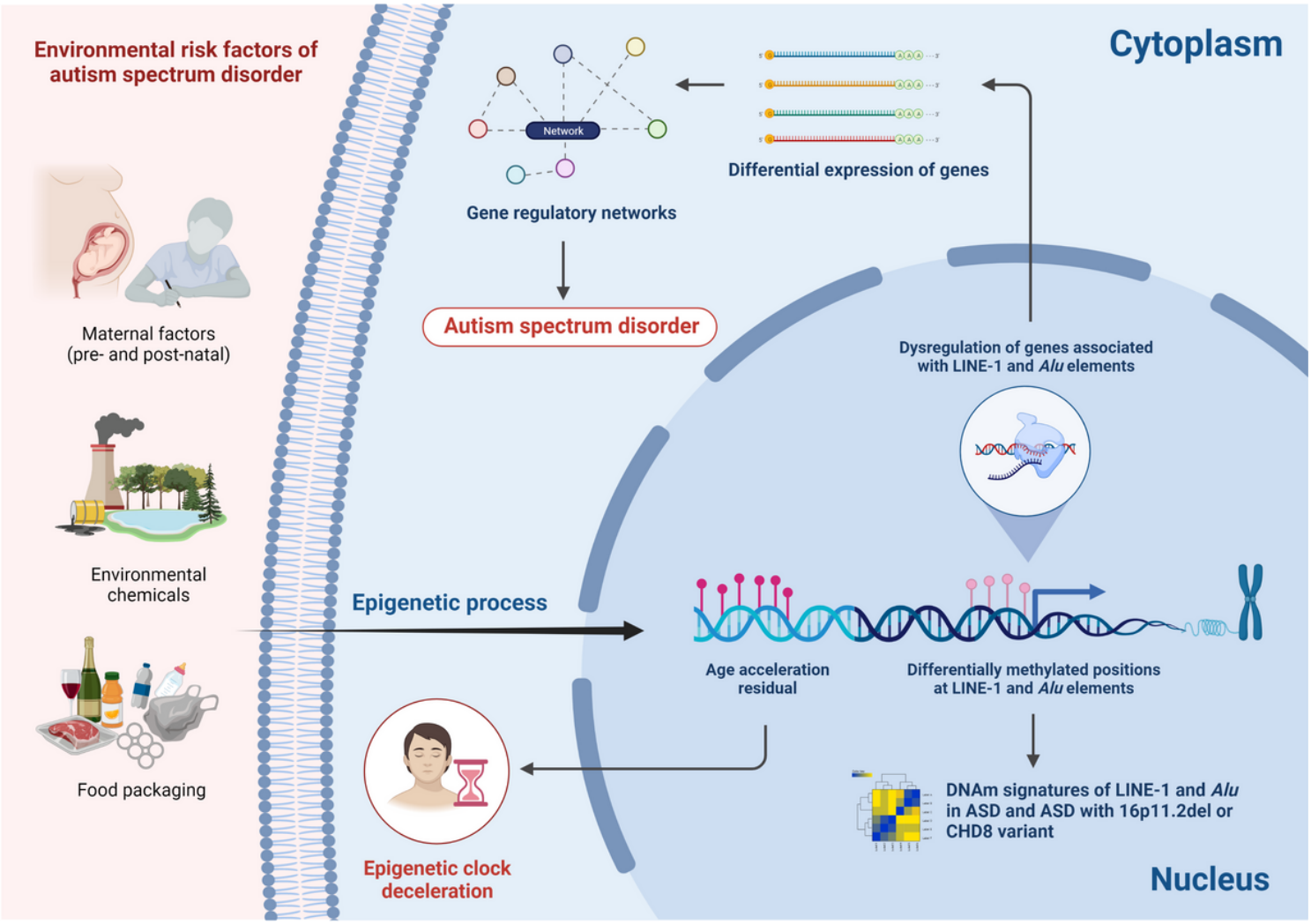


Figure 10

Schematic diagram illustrating a possible mechanism of LINE-1 and *Alu* elements in ASD.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.docx
- Additionalfile2.tif
- Additionalfile3.tif
- Additionalfile4.tif
- Additionalfile5.docx
- Additionalfile6.docx
- Additionalfile7.docx
- Additionalfile8.tif
- Additionalfile9.tif
- Additionalfile10.tif