Brain Research 1724 (2019) 146457

Contents lists available at ScienceDirect

Brain Research

journal homepage: www.elsevier.com/locate/brainres

Research report

Artificial intelligence analysis of newborn leucocyte epigenomic markers for the prediction of autism

Ray O. Bahado-Singh[a], Sangeetha Vishweswaraiah[a], Buket Aydas[b], Nitish K. Mishra[c], Ali Yilmaz[a], Chittibabu Guda[c], Uppala Radhakrishna[a,*]

[a] Department of Obstetrics and Gynecology, Oakland University William Beaumont School of Medicine, Royal Oak, MI, USA
[b] Department of Mathematics & Computer Science, Albion College, Albion, MI, USA
[c] Department of Genetics, Cell Biology & Anatomy, College of Medicine, University of Nebraska Medical Center, Omaha, NE, USA

HIGHLIGHTS

- We studied the changes of DNA methylation present in newborn leucocyte DNA.
- We identified significant methylation changes in numerous genes linked to brain function and autism.
- Early diagnosis and intervention will improve disease outcomes.
- Artificial Intelligence using methylation data identified accurate biomarker algorithms for autism prediction.

ARTICLE INFO

Keywords:
Autism
DNA methylation
Epigenetics
Artificial intelligence
Newborns

ABSTRACT

A great diversity of factors contribute to the pathogenesis of autism and autism spectrum disorder (ASD). Early detection is known to correlate with improved long term outcomes. There is therefore intense scientific interest in the pathogenesis of and early prediction of autism. Recent reports suggest that epigenetic alterations may play a vital role in disease pathophysiology. We conducted an epigenome-wide analysis of newborn leucocyte (blood spot) DNA in autism as defined at the time of sample collection. Our goal was to investigate the epigenetic basis of autism and identification of early biomarkers for disease prediction. Infinium HumanMethylation450 BeadChip assay was performed to measure DNA methylation level in 14 autism cases and 10 controls. The accuracy of cytosine methylation for autism detection using six different Machine Learning/Artificial Intelligence (AI) approaches including Deep-Learning (DL) was determined. Ingenuity Pathway Analysis (IPA) was further used to interrogate autism pathogenesis by identifying over-represented biological pathways. We found highly significant dysregulation of CpG methylation in 230 loci (249 genes). DL yielded an AUC (95% CI) = 1.00 (0.80–1.00) with 97.5% sensitivity and 100.0% specificity for autism detection. Epigenetic dysregulation was identified in several important candidate genes including some previously linked to autism development e.g.: *EIF4E, FYN, SHANK1, VIM, LMX1B, GABRB1, SDHAP3* and *PACS2*. We observed significant enrichment of molecular pathways involved in neuroinflammation signaling, synaptic long term potentiation, serotonin degradation, mTOR signaling and signaling by Rho-Family GTPases. Our findings suggest significant epigenetic role in autism development and epigenetic markers appeared highly accurate for newborn prediction.

1. Introduction

Autism spectrum disorder (ASD) is a neuro-developmental disorder which is complex and characterized by impaired communication difficulties, social interactions and repetitive or compulsive behaviors that generally begin within the first three years of life (Xiao et al., 2014). Symptoms and findings may also include sensory processing difficulties, anxiety or irritability, pica, sleep dysfunction, seizures or gastrointestinal disorders (Coury et al., 2012). Around 75% children with ASD have lifelong disability requiring substantial social and educational support (Mefford et al., 2012). Boys are more prone to ASD than girls with a 4:1 to 5:1 ratio (American Psychiatric Association, 2013). In

* Corresponding author at: Department of Obstetrics and Gynecology, Oakland University William Beaumont School of Medicine, 3601 West 13 Mile Road, Royal Oak, MI 48073, USA.
E-mail address: Uppala.radhakrishna@beaumont.edu (U. Radhakrishna).

https://doi.org/10.1016/j.brainres.2019.146457
Received 2 July 2019; Received in revised form 10 September 2019; Accepted 11 September 2019
Available online 12 September 2019
0006-8993/

1994 the American Psychiatric Association (DSM-IV) recognized multiple disease subtypes: autistic disorder, Rhett syndrome, Asperger syndrome, childhood disintegrative disorder (1994). In 2014 however in the DSM-5, these disorders were grouped into a single diagnosis "Autism Spectrum Disorder" (American Psychiatric Association, 2013; Vahia, 2013).

ASD is one among the fastest-increasing developmental disorders in the U.S (McCarthy, 2014). While changes in the definition and reduced frequency of under-diagnosis may partly contribute to this rise, there is evidence that greater than 70% of the reported increase in USA cases is in fact due to a true increase in disease prevalence (Nevison, 2014).

A full understanding of the disease pathogenesis continues to elude researchers. ASD is known to be highly heterogeneous in both clinical presentation and etiology. There is evidence of a strong genetic component (Mendelsohn and Schaefer, 2008; Betancur, 2011). Classic genetic changes, such as gene mutations, translocations, inversions, and copy number variations play a major role in the genetic etiology of ASD (Mefford et al., 2012; Reddy, 2005). Multiple ASD susceptibility genes have been identified in the past decade. Several genetic syndromes are associated with ASD including: Angelman, Asperger's and Fragile X and others. Further, copy number variations are found with greater frequency in individuals with ASD. Collectively however, genetic factors only account for approximately 10–20% of ASD cases.

Conservatively therefore in at least 70% of ASD cases the underlying mechanisms do not appear to be genetic and remain largely unknown (Sanders et al., 2011). Non-genetic mechanisms implicated in ASD pathogenesis, include environmental and immunological factors with significant data suggesting a significant contribution by prenatal and newborn exposures (Grice and Buxbaum, 2006; Matelski and Van de Water, 2016; Sealey et al., 2016).

Epigenetics refers to alterations in gene function not due to mutations. The expression levels of a gene are largely influenced by the epigenetic state of its surrounding regulatory regions. A range of environmental factors influence gene expression through epigenetic mechanisms. Current evidence suggests that epigenetic mechanisms could play a significant role in ASD development (Hu, 2013; Loke et al., 2015). The Methyl-CPG-binding protein 2 (*MECP2*) gene for example provides interesting insight into the potential role of epigenetic mechanisms in ASD. The *MECP2* gene locus is on the X-chromosome. The protein binds to methylated regions of DNA and in turn other gene transcription suppressors bind to *MECP2* and inhibit transcription. While being involved in the epigenetic control of transcription process genome-wide, *MECP* also plays a critical role in synaptogenesis and long-term synaptic plasticity (Kavalali et al., 2011). Aberrant *MECP2* promoter methylation (Nagarajan et al., 2006) is associated with a significant reduction of *MECP2* protein expression. *MECP2* mutations have been reported in autism (Lopez-Rangel and Lewis, 2006) however these are rare.

The American Academy of Pediatrics recommends that all children be screened between 18 and 24 months for ASD, with subsequent ongoing developmental monitoring (Johnson et al., 2007). This recommendation is supported by evidence that early interventions can reduce deficits in including, social attention, language, IQ, and overall symptom severity (Dawson et al., 2010). Currently however, large areas of the USA do not perform the recommended early clinical screening. The development of accurate disease biomarkers would enable early disease detection and facilitate interventions to improve outcomes.

Artificial Intelligence (AI) holds promise as an analytic tool for identifying patterns and associations that can transform biological big data such as epigenomics into accurate biomarkers for complex disorders (Bahado-Singh et al., 2019a). Machine Learning (ML) is a branch of AI. We therefore evaluated and compared the performance of six ML/AI techniques including Deep Learning (DL), the latest branch of ML, for the prediction of ASD. DL is a representation learning algorithm. It mimics the functioning of the human brain. Multiple artificial 'neuronal' layers are utilized between the input and output layers in order to generate automated classifications and predictions by the computer program. Nowadays, DL is widely used in image processing, pattern recognition and recently in healthcare such as for the extraction of relevant patterns from electronic health records. A deep neural network (DNN) is an artificial neural network (ANN) with multiple layers between the input and output layers (Bengio, 2009; Schmidhuber, 2015). The DNN finds the correct mathematical manipulation to turn the input into the output, whether it is a linear relationship or a non-linear relationship. DL models are inspired by information processing and communication patterns in biological nervous systems. The network moves through the layers calculating the probability of each output.

The existence of prenatal and newborn risk factors in ASD development suggests the feasibility of newborn screening for this disorder. We therefore looked at the use of DL and other ML techniques for the detection of autistic disorder (the classification system then used at the time of sample generation) based on DNA methylation status of newborn leucocyte DNA. In addition, based on cytosine ('CpG' or cg') loci (and related genes) that were significantly epigenetically altered, we investigated that potential epigenetic pathogenesis of autism.

2. Results

The clinical data for cases and controls are shown in Supplemental Table S1. There were no statistically significant differences between the two groups and no batch effect was identified in the laboratory analysis. We identified a total of 230 CpG sites associated with 249 distinct genes that were significantly differentially methylated between ASD cases and the control [false discovery rate (FDR) ≤ 0.05]. These included coding genes, open reading frame genes, LOC (uncharacterized) genes and small nuclear RNA genes. The list of Top 25 significant CpG sites and their relevant gene and chromosome locations are shown in Table 1, and the remaining significant CpG sites are listed in Supplemental Table S2. These CpG sites include 160 that were significantly differentially hypomethylated and 70 that were hypermethylated. There were 87 CpGs that met the stringent genome wide threshold i.e. p-value of $< 5 \times 10^{-8}$. The Heat Map identified gene clusters (clusters 1–3) that provided good visual separation of autism cases from controls (Fig. 1).

Among individual cytosine loci (and corresponding genes), the best predictive individual CpG markers (genes) i.e. AUC $\geqq 0.9$ for autism detection were: cg20129082 (*LOC100126784*; *NAV2*), cg08590939 (*OXCT1*), cg11722376 (*LOC389033*), cg15371711 (*MYL9*), cg16678169 (*ALS2CR4*), cg15028160 (*C19orf73*) and cg01156550 (*ASCL2*). The 4 example AUC-ROC curves are presented in Fig. 2. All of these CpG loci had $\geq 5\%$ methylation difference between autism cases and controls- a threshold methylation difference that likely increases the biological significance i.e., the observed methylation changes are likely to be associated with actual changes in gene expression (Leenen et al., 2016). A total of 101 distinct CpGs had individual AUC ≥ 0.75 (a threshold suggestive of biological significance i.e. change in gene expression) with the FDR p-value < 0.05, for distinguishing autism from controls, Table 1 and Supplemental Table S2. On differentially methylated region (DMR) analysis, we identified one region on chromosome 11: 2293048–2293117 comprising 3 CpGs with a FDR p = 1.61E-32.

We next performed a search for the genes that were differentially expressed in the brain of autistic cases. We used previously published transcriptomic data (Voineagu et al., 2011). We could identify 15 genes that were differentially methylated in leukocyte DNA in our study and which were also reported to be differentially expressed in publication of Voineagu et al. (2011). These genes are presented in Table 2. We subsequently used the methylation level of the CpG loci of these genes for autism detection. When we matched our differentially methylated genes to Genotype-Tissue Expression (GTEx) project data with brain cerebellum tissue, we found that the *APLP2* and *ATP5B* genes were also differentially expressed in cerebellum region.

Table 1
Top 25 differentially methylated genes with Target ID, Gene ID, chromosome location, % methylation change and FDR p-value for each gene methylated. Each CpG locus displayed significant methylation change FDR p-value < 0.05 with an individual AUC ≥ 0.75 for Autism detection.

Target ID	Chr	Genes	FDR p-Val	Fold change	% Methylation		% of Methylation difference	AUC	CI	
					Cases	Controls			lower	upper
cg09167828	2	C2orf50	9.545E-56	1.96	76.27	38.90	37.54	0.79	0.60	0.99
cg26069044	11	NRGN	8.999E-55	2.18	67.06	30.82	36.44	0.76	0.56	0.97
cg19935756	4	GABRB1	5.025E-53	2.35	59.29	25.18	34.31	0.79	0.60	0.99
cg20187719	3	LOC285375	8.095E-51	5.87	37.62	6.41	31.21	0.79	0.60	0.99
cg07319199	2	ANTXR1	2.219E-49	1.55	82.04	52.87	29.29	0.71	0.49	0.93
cg10989317	10	UBTD1	5.839E-49	1.51	84.62	56.07	28.61	0.86	0.70	1.00
cg06523151	1	ESRRG	2.054E-47	1.50	78.15	51.96	26.49	0.75	0.54	0.96
cg09247979	6	PTPRK	3.365E-47	1.75	60.34	34.49	26.00	0.69	0.47	0.91
cg03979311	5	GZMK	8.129E-47	1.72	60.16	34.93	25.37	0.74	0.53	0.95
cg07876831	13	TMCO3	1.486E-46	1.50	74.05	49.24	25.10	0.67	0.45	0.90
cg12307373	11	DYNC2H1	2.528E-46	1.84	53.67	29.24	24.59	0.74	0.53	0.95
cg20592836	20	TP53INP2	2.374E-45	2.33	39.85	17.07	22.77	0.73	0.52	0.94
cg20089799	12	TSPAN9	2.583E-45	1.89	48.34	25.63	22.76	0.70	0.48	0.92
cg04787784	16	LMF1	4.541E-45	1.58	60.52	38.23	22.31	0.69	0.46	0.91
cg11973981	7	ISPD	1.799E-44	1.53	61.25	40.03	21.71	0.71	0.50	0.93
cg01127608	9	LMX1B	2.789E-44	2.75	32.82	11.94	20.86	0.66	0.43	0.89
cg08477332	1	S100A14	1.136E-43	1.66	49.90	30.14	19.80	0.68	0.45	0.90
cg11585022	5	WDR36	3.013E-43	1.56	52.68	33.72	19.00	0.65	0.42	0.88
cg08407901	21	SLC37A1	8.170E-43	2.17	33.62	15.50	18.12	0.64	0.40	0.87
cg06417478	19	HOOK2	2.502E-42	1.74	40.43	23.27	17.16	0.63	0.40	0.86
cg16678169	2	ALS2CR4	1.211E-28	0.14	3.83	26.67	−22.88	0.96	0.88	1.00
cg19746536	19	PPP1R15A	2.337E-28	0.17	5.10	29.22	−23.98	0.84	0.66	1.00
cg05455372	7	LOC168474	4.814E-28	0.63	43.01	68.55	−25.42	0.81	0.63	1.00
cg11144103	17	PTRF	5.790E-28	0.61	41.04	66.94	−25.83	0.78	0.58	0.98
cg09084244	12	CDK2AP1	6.639E-28	0.61	41.62	67.80	−26.14	0.75	0.54	0.96

2.1. Artificial Intelligence analysis results

We evaluated several models of machine learning tools to identify best predictive markers and algorithms for autism. ML techniques appeared highly accurate for the prediction of autism. DL achieved AUC values of 0.958–1.00 for the detection of autism using a combination of 5 CpG loci. Similarly, the AUCs for autism detection was ≧0.95 for four of the other 5 ML approaches. The values of the different AI platforms in the different CpG (gene) groups are as provided: (i) analysis evaluating the top 230 CpG markers (Table 3), (ii) based on CpG markers meeting the stringent p-value threshold ($< 5 \times 10^{-8}$) (Supplemental Table S3) and (iii) CpG loci from the 15 differentially expressed genes in autism cases brain (Supplemental Table S4) that also were found to be differentially methylated in newborn leukocytes. DL achieved an AUC (95% CI) = 1.0, sensitivity 97.5% and specificity 100% for ASD detection when only high stringency markers ($p < 5 \times 10^{-8}$) were considered (Supplemental Table S3).

2.2. Standard multivariate logistic regression analysis

The combination of 3 CpG/gene markers: cg20129082 (*LOC100126784; NAV2*), cg08590939 (*OXCT1*) and cg20187719 (*LOC285375*) achieved an AUC 95% CI = 1.0 (1.0, 1.0). This confirms the robustness of CpG markers for autism prediction.

2.3. Gene functional and pathway enrichment

We performed molecular pathway and disease enrichment analysis using the differentially methylated genes to further understand the mechanisms of autism. We observed significant enrichment of genes involved in neuroinflammation signaling, synaptic long term potentiation, serotonin degradation, mTOR signaling and signaling by Rho Family GTPases (Supplemental Fig. S1). The roles of the molecular pathways in neurological function is presented in Supplemental Table S5. Genes such as *EIF4E, FYN, SHANK1, VIM, LMX1B, GABRB1, SDHAP3* and *PACS2* that were found to be differentially methylated in our study have previously been identified as having a role in brain function (Supplemental Table S6).

3. Discussion

In the present analysis we have identified 230 significantly differentially methylated CpG loci in 249 genes in leucocyte DNA of newborns subsequently diagnosed with autistic disorder. Using multiple Artificial Intelligence AI/ML platforms and epigenomic data we were able to accurately predict autism. Deep Learning and four of the five other ML approaches achieved AUCs ≧0.95. Children who are diagnosed early and who participate in intervention programs are reported to have better long term outcomes (Zwaigenbaum et al., 2015). Screening starting at 16 months and on an ongoing basis has therefore been recommended (Armstrong, 2008). Currently, there are no validated preclinical biomarkers of ASD. As noted earlier, The American Academy of Pediatrics has recommended that all children be screened for ASD between 18 and 24 months with subsequent ongoing developmental monitoring (Johnson et al., 2007) and there is supporting evidence that early interventions can reduce deficits of ASD including social attention, language, IQ, and the severity of symptoms (Dawson et al., 2010; Kasari et al., 2010). The uptake of routine screening has however been less than desired in many areas of the US (Zwaigenbaum et al., 2015). Accurate predictive biomarkers could have clinical utility. The timing of the analysis- within days of birth is therefore potentially significant given the above evidence and guidelines.

The generation of mechanistic data was an important objective of our study. We identified multiple CpG loci and corresponding genes that displayed significant methylation changes in autism cases. Some of these genes have previously been linked or can plausibly be linked to ASD. A sub-set is detailed in Supplemental Table S6. For most of the differentially methylated genes that we have identified no significant association with autism or the expanded ASD has yet been reported to our knowledge. Further work will need to be done in this area. We briefly discuss a few of the genes that were differentially methylated in our study that have or can be credibly linked to autism. The role of LIM Homeobox Transcription Factor 1 Beta (LMX1B) in the development and maintenance of serotoninergic neurons is well known. Thanseem

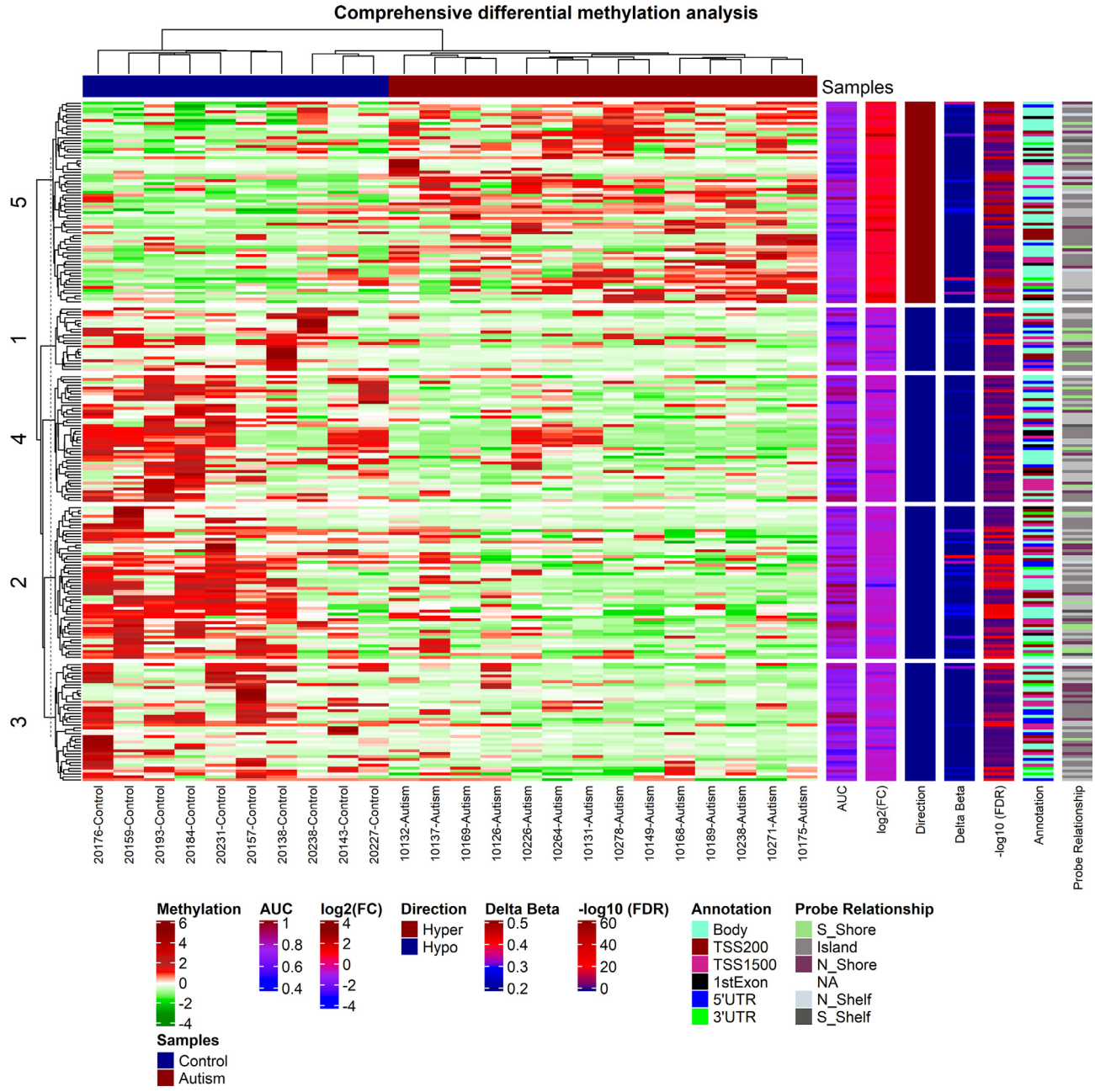

Fig. 1. Heatmap of 28 highly differentially methylated loci between autism cases and controls. DNA methylation profiling based on unsupervised hierarchical clustering identified five unique clusters having distinct methylation signatures. The figure also displays direction, fold change in disease, probe relationship and probe annotation of differentially methylated CpG sites.

et al. (2011) conducted LMX1B gene mRNA expressions studies on postmortem brain tissues of autism cases versus unaffected controls and identified significantly lower expression in the anterior cingulate gyrus region of autism patients. They suggested a possible role of LMX1B in the pathophysiology of autism. In another study Bergman et al. (2010) found an association between LMX1B and schizophrenia. We observed hypermethylation in this gene with a 20% difference in methylation levels between cases and controls. A study by Ladd-Acosta et al. (2014) using post-mortem brain tissue from autism cases identified differentially methylated regions on SDHAP3 gene. In our study, we found a significant association between a transcription start site CpG cg08422420 (SDHAP3) hypermethylation and autism. A study by Krumm et al. (2013) reported an inherited copy number variation affecting PACS2 gene in families affected by autism spectrum disorder. In the present study we found a significantly differentially methylated CpG locus (cg12425861) within the PACS2 gene in the autism cases.

3.1. Molecular pathways in autism

The molecular pathways that are enriched in our analysis are listed in (Supplemental Table S5). Neuroinflammation signaling is one such pathway. Neuroinflammation signaling pathways were previously reported to be altered in ASD development (Dipasquale et al., 2017). Studies have reported microglial activation with elevated pro-inflammatory cytokines and chemokines in brain regions such as the frontal cortex, the cerebellum and in the cerebrospinal fluid of autism cases (Nakagawa and Chiba, 2016). These suggest an important role of inflammation in autism.

Signaling by Rho family GTPases plays an important role in brain development (Narumiya and Thumkeo, 2018). During the neuronal

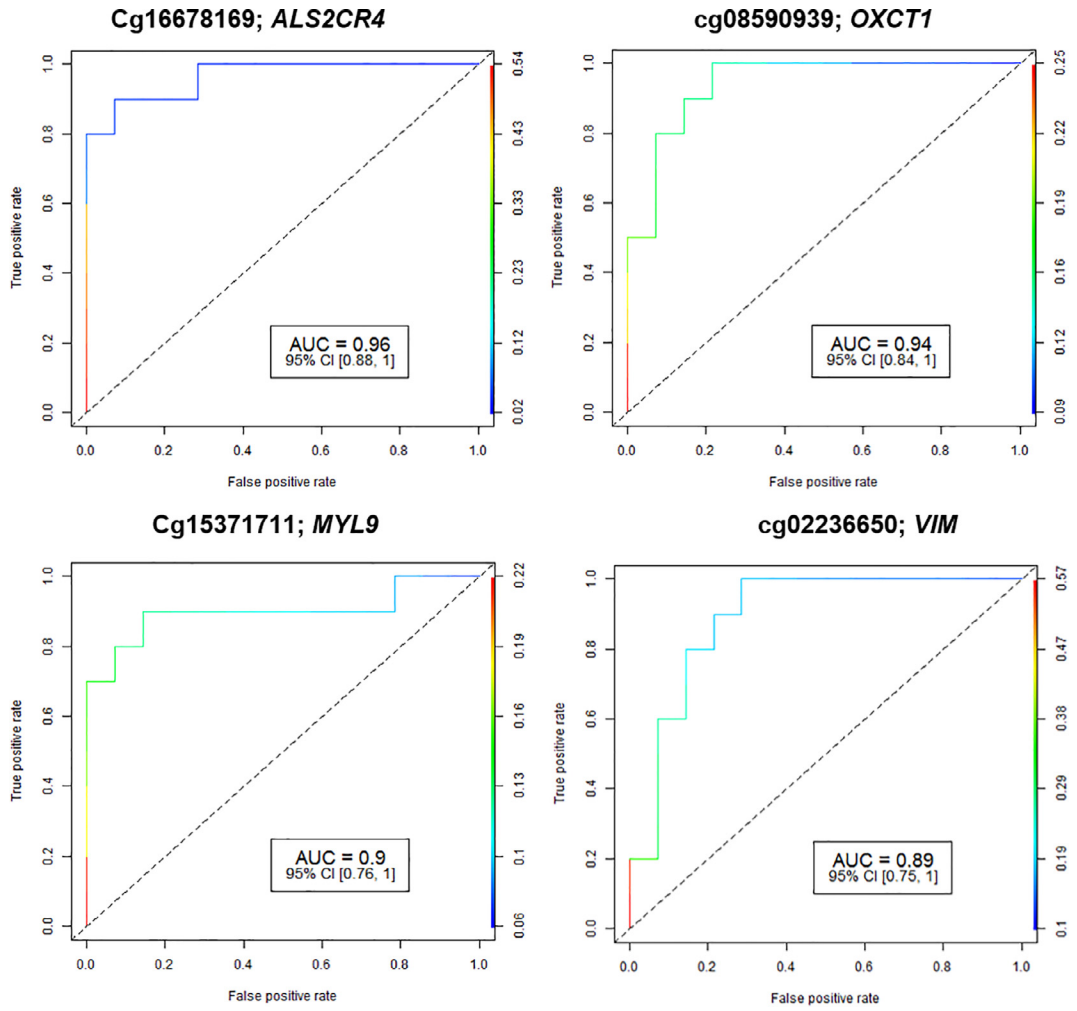

Fig. 2. Receiver operating characteristic (ROC) curves of methylation profiles for four specific markers associated with Autism.

development, cadherin family proteins regulate the Rho family GTPases (Govek et al., 2005). We identified two cadherins *CDH15* and *CDH3* in the Rho family GTPases signaling pathway that were significantly differentially methylated. Impaired cadherin function leads to loss in the functional connectivity and coherent information processing in the brain (Redies et al., 2012). One of the identified cadherin genes *CDH15* has been linked to cognitive impairment and autism (Redies et al., 2012). We also found epigenetic dysregulation of the *PCDH8* gene. The gene product is a protocadherin from the cadherin superfamily. This is a putative causal gene for autism (Butler et al., 2015). We also identified genes enriched in mTOR signaling in our study. mTOR signaling is needed for neuronal and glial differentiation followed by the maintenance of the stemness of neural stem cells during brain development (Lee, 2015). mTOR signaling regulates synaptic plasticity and

Table 2
Differentially methylated leucocyte genes found also to be differentially expressed in brain samples of ASD cases from Voineagu et al. study.

Target ID	Chr	Genes	FDR p-Val	Fold change	% Methylation		% of Methylation difference	AUC	CI	
					Cases	Controls			lower	upper
cg02236650	10	VIM	4.20274E-07	0.62	17.37	27.93	−10.68	0.89	0.75	1.00
cg18339359	8	SLC25A37	3.25651E-25	0.57	52.23	92.28	−40.57	0.86	0.69	1.00
cg00613562	8	SULF1	1.22503E-05	0.62	13.70	22.21	−8.55	0.81	0.63	1.00
cg06523151	1	ESRRG	2.05437E-47	1.50	78.15	51.96	26.49	0.75	0.54	0.96
cg03748376	1	OR2L13	1.26758E-15	2.00	24.65	12.33	12.31	0.71	0.50	0.93
cg07319199	2	ANTXR1	2.2189E-49	1.55	82.04	52.87	29.29	0.71	0.49	0.93
cg02750262	18	ZADH2	3.99185E-07	0.57	12.62	22.08	−9.49	0.71	0.49	0.93
cg20089799	12	TSPAN9	2.58266E-45	1.89	48.34	25.63	22.76	0.70	0.48	0.92
cg12743416	7	TRIM24	9.70041E-16	0.51	17.45	34.35	−16.88	0.69	0.47	0.91
cg24791218	4	FSTL5	1.68543E-07	0.30	2.78	9.27	−6.49	0.69	0.47	0.91
cg16875032	8	SNTB1	0.00243946	1.50	15.91	10.58	5.30	0.66	0.44	0.89
cg18418335	19	MKNK2	0.045426808	0.45	2.42	5.40	−3.03	0.66	0.43	0.89
cg11478607	22	GSTT1	0.000107618	1.65	15.47	9.39	6.39	0.57	0.33	0.81
cg24668570	10	KNDC1	0.001128763	0.58	7.61	13.05	−5.48	0.52	0.28	0.76
cg13163765	6	SLC25A27	0.007377812	0.61	7.40	12.07	−4.70	0.75	0.54	0.96

Table 3
Results of autism prediction (based on to 230 markers).

	SVM	GLM	PAM	RF	LDA	DL
AUC	0.9642	0.9791	0.9500	0.9857	0.9583	0.9791
95% CI	(0.7642–1)	(0.7791–1)	(0.7500–1)	(0.7857–1)	(0.7583–1)	(0.7791–1)
Sensitivity	1.0000	0.9000	0.8750	0.8750	0.8500	1.0000
Specificity	0.8500	0.9500	0.9500	0.9750	0.9800	0.9000

Important predictors per AI platform in order of decreasing contribution:
SVM: cg16678169, cg20129082, cg06731443, cg08590939, cg08634464.
GLM: cg01156550, cg20129082, cg08590939, cg15371711, cg02648941.
PAM: cg16678169, cg26846943, cg11701583, cg02502145, cg05859760.
RF: cg08590939, cg24361896, cg20129082, cg11380830, cg02313495.
LDA: cg20129082, cg16678169, cg08590939, cg22772691, cg05317207.
DL: cg01156550, cg11480019, cg17197981, cg24506221, cg25831233.

neurodevelopment and when dysregulated, leads to the development of autism spectrum disorder (Onore et al., 2017; Winden et al., 2018).

Our data has some limitations. We were unable to perform expression analysis given that we used dried blood spots. We however sought to verify whether any of the genes previously reported to be differentially expressed in autism brain (Voineagu et al., 2011) were also differentially methylated in leucocyte DNA. We found 15 such genes. Among them, 13 genes were found to be differentially methylated by more than 5%, compared to controls a level of methylation difference that suggests differential gene expression. Among the 15 genes, *SLC25A37* and *SLC25A27* were related with mitochondrial function and autism (Anitha et al., 2012). The value of the development of leucocyte epigenetic markers is not primarily to identify blood markers with the exact expression or DNA methylation patterns as brain tissue. Indeed there is likely to be significant variation in the different regions of the same brain in an affected individual. Rather the objective was to identify epigenetic markers in leucocytes that correlate with (positively or negatively) genomic changes in the brain and can serve as markers of or signposts of changes occurring in the brain.

Another limitation was that the blood spots were obtained and archived in a period when an earlier classification system of autism was being used. We limited our analysis to autism (current classification Autistic disorder/autism) and not ASD. Our findings therefore might not apply to other current subtypes of ASD such as Asperger and Pervasive Developmental Disorder. Given the relatively small numbers of cases in this proof of concept study, further analyses of larger number of cases to confirm our findings would be highly desirable. Dispite the relatively small sample size, we did however achieve highly accurate prediction. It is possible however that with larger study numbers additional CpG loci/genes might have been found to have undergone statistically significant epigenetic modification and thus provide more complete mechanistic information on autism.

3.2. Conclusion

In this proof of concept study, using easily accessible blood leucocyte DNA, we were able to evaluate epigenetic changes associated with and potentially causally linked to autism development. Our data provided supportive evidence of a significant role of epigenetic modification in autism pathogenesis. By combining multiple AI platforms and leucocyte epigenomics we were able to achieve consistently high accuracy for autism prediction in the newborn period. Despite the clear importance of early prediction, currently there are no early biomarkers in clinical use making our findings potentially more consequential. Given the inaccessibility of brain tissue in living cases for analysis, the use of easily accessible leucocyte DNA markers for ASD assessment and prediction could represent a significant scientific and clinical advance. We recommend larger follow up studies to validate our findings.

4. Methods and materials

Neonatal dried blood spots were used to extract genomic DNA using QIAamp® (Qiagen) DNA extraction kit as per the manufacturer's protocol. DNA from archived dried blood spot (DBS) serves as templates and previously been used for genome-wide DNA methylation profiling using the Infinium HumanMethylation450 BeadChip assay (Joo et al., 2013; Radhakrishna et al., 2016; Wong et al., 2008). For the purpose of mandated newborn screening treatment program, blood spot specimens were collected by the Michigan Department of Community Health in the State of Michigan (MDCH) now the Michigan Department of Health and Human Services (MDHH). All specimen collection was performed between 24 and 79 h after birth. Parent(s) or a legal guardians had consented to allow their child's remaining blood for research purpose after the clinical testing. The present study was approved by the Wayne State University Institutional Review Board and by the the Michigan Department of Community Health Institutional Review Board. Since cases were diagnosed prior to 2014, i.e. between 2000 and 2010, the diagnostic category used was autistic disorder subtype as per the then current American Psychiatric Association (DSM-IV) classification prior to 2014. Cases were diagnosed and followed by neurologists with the hospital medical records confirming this diagnosis. Participants were excluded if they had any known or suspected genetic syndromes or other major congenital anomalies. A limited demographic data such as date of sample collection, newborn sex, maternal race and age followed by gestational age at delivery was collected. Patients with autism variants and not having autistic disorder as per the definition at the time of sample collection were excluded from the study group. Unaffected controls had no major medical, surgical disorders or known or suspected major congenital anomaly or suspected syndromic disorders on chart review. It should be noted that chart review for cases and controls were performed several years after sample collection providing adequate time for major medical disorders to be identified. Our study cohort included 14 autism cases and 10 normal participants. Demographic and clinical characteristics during the newborn period were compared (Supplemental Table S1).

4.1. Genome-wide methylation analysis using the HumanMethylation450

Methylation analysis of whole genome was performed for 24 participants (14 ASD and 10 control participants) using the HumanMethylation450, Illumina's Infinium® HD BeadChip assay (Illumina, Inc., California, USA). All samples were randomized and processed at once in a single plate to avoid batch effect. The BeadChip contains probes for 485,577 CpG methylation sites and 500 ng of genomic DNA was used. These CpG sites represent 96% of RefSeq genes, 95% of CpG islands and approximately 17 CpG sites per covered gene region including the promoter (TSS200 and TSS1500), 5′UTR, 3′UTR, coding and gene body regions. Methylation profiling of DNA using Illumina Infinium assay with peripheral blood leucocytes has

been used previously to detect CpG sites associated with disease phenotypes (Bahado-Singh et al., 2019a; Docherty et al., 2014; Pan et al., 2012; Radhakrishna et al., 2018).

The laboratory method has been previously described (Radhakrishna et al., 2018). We used EZ DNA Methylation Kit (Zymo Research, Orange, CA) for bisulfite conversion of DNA per the manufacturer's instructions followed by methylation array processing. Case and control samples were randomized on the beadchips to avoid batch effect. The fluorescently stained methylation BeadChips were imaged with the Illumina iScan. Prior to detailed bioinformatic and statistical analysis, data preprocessing and quality control were performed. The raw "idat" files from cases and control were normalized. These included examination of the background signal intensity of both cases and controls, methylated and unmethylated signals, and the ratio of methylated to unmethylated signal intensities. The processing was performed per the manufacturer's protocol and 99% of the methylation sites were unequivocally determined.

4.2. Statistical and bioinformatic analysis

Genome-wide DNA methylation sites were analyzed using the "GenomeStudio" (Illumina) analysis package. Following the pre-processing, DNA methylation ß-value, a measure of percentage methylation, was calculated for each CpG site. CpG sites around the SNPs with a distance of 10 bp and CpGs associated with X and Y chromosomes were excluded for the downstream analysis to avoid potential confounding factors and/or gender bias (Chen et al., 2013; Liu et al., 2013; Wilhelm-Benartzi et al., 2013). It has been showed that the CpGs associated with SNPs within the probe sequence (10 bp) influences the corresponding methylated probes (Daca-Roszak et al., 2015). The remaining methylation sites were considered for further analysis. We have performed Differentially methylated region (DMR) analysis to ascertain the methylation status across different biological samples using "DMRcate" package of R.

4.3. Expression analysis

We also performed a subanalysis, we mined the data to identify genes previously reported to be differential expressed in brain samples of autism cases (Voineagu et al., 2011) that were also differentially methylated in our leucocyte study. We subsequently determined the performance of autism prediction using AI techniques for analysis of the methylation results. We also correlated our differentially methylated genes to Genotype-Tissue Expression (GTEx) project data with brain cerebellum tissue.

4.4. Artificial intelligence/machine learning analysis

4.4.1. Data set

There are three groups of data that were evaluated in this study. First one consisted of CpG markers (each associated with separate genes) that individually (each CpG) had an AUC Benjamini-Hochberg False Discovery Rate (FDR p-value < 0.05) for autism prediction based on its methylation level. The second group of predictors used individual CpG's with highly significant methylation differences in autism cases versus controls (p-value $< 5 \times 10^{-8}$). This stringent p-value threshold has been recommended for genome-wide association studies to define genome-wide threshold of significance (Risch and Merikangas, 1996). This threshold minimizes false-positives while minimizing the chances of missing scientifically important associations. Finally, as noted previously, we also looked at the performance of CpG loci in genes in our study that were previously reported by Voineagu and collaborators (Voineagu et al., 2011) to be differentially expressed in the brain of autism cases.

4.4.2. Data preparation

Missing values were detected and replaced by a value which represents half of the lowest positive values in the original data. The log value of each predictor was adjusted by its mean and auto scaled by its standard deviation. Quantile normalization techniques were used to minimize sample-to-sample difference.

4.4.3. Artificial intelligence algorithms

Artificial intelligence techniques are utilized to make predictions directly from a set of features that have been pre-labeled by the user i.e. classical machine learning algorithms, or to map features into a transitional representation before transforming them to ultimately generate predictions i.e. representation learning algorithms (Wang, 2017).

In addition to DL which is a representation learning algorithm, five classical machine learning algorithms were also used in this study. We compared the predictive accuracy: AUC, specificity, and sensitivity for all six AI algorithms. We selected these five machine learning algorithms since their specific properties are appropriate for analysis of metabolomic and genomic data. The five additional ML algorithms used were summarized as follows. Random forest (RF) is based on decision tree theory and is used for high-dimensional data. The difference between standard trees and Random Forest is based on the split process in which the standard trees choose the best split while in Random Forest the splits are chosen randomly. This strategy performs very well compared to many other classifiers and is robust against overfitting. Support vector machine (SVM) has the advantage of avoiding overfitting and uses the kernel trick for more complex problems - for cases in which no linear separation is possible, and achieves better results by altering the kernel function. Generalized Linear Model (GLM) identifies the linear connection between the predictors and the dependent variable. Since it is a linear algorithm, the output of a GLM is informative in terms of the relationship between dependent and independent variables. Prediction Analysis for Microarrays (PAM) is used for the classification of gene expression data using the nearest shrunken centroids and gives satisfying results in metabolomic and genomic studies. Linear Discriminant Analysis (LDA) attempts to explain one dependent variable as a linear combination of predictors. The AI algorithms used in this study have been previously used in metabolomics and genomics studies (Alakwaa et al., 2018; Alpay Savasan et al., 2019; Bahado-Singh et al., 2018, 2019a,b).

4.4.4. Software tools

We utilized the h2o R package (https://cran.r-project.org/web/packages/h2o/h2o.pdf, Author: The H2O.ai team Maintainer Tom Kraljevic < tomk@0xdata.com >) to modify the parameters of the DL algorithm and the caret R package (https://cran.r-project.org/web/packages/caret/caret.pdf, Maintainer Max Kuhn < mxkuhn@gmail.com > , December 10, 2017) to adjust the parameters in the other AI algorithms.

The predictor ranking functions *varimp* in h2o and *varImp* in caret R packages were utilized to rank or order the predictors in each of the models.

We used the pROC R package to compute the area under the ROC curve (AUC), and specificity and sensitivity of all the different AI models.

4.4.5. Modeling & evaluation

The data were separated into training and validation or testing sets. We preferred to split the data set as follows: 80% training and 20% testing, as this ratio is generally used in medium sized data sets. 10-fold cross validation analysis was performed on the training group of the data during the model formation stage. In addition, the cross validation analysis was performed ten times and the average AUC, sensitivity, specificity and 95% confidence intervals for the test set were calculated.

Several parameters were used to tune the models namely: adjustment of the number of trees for RF, classification cost for SVM and

threshold level for shrinking toward the centroid for PAM. For DL, model adjustments similarly made were: a) *Epochs* (number of passes of the full training set), b) *l1* (penalty to converge the weights of the model to 0), c) *l2* (penalty to avoid the expansion of the weights), d) *input_dropout_ratio* (ratio of disregarded neurons in the input layer), e) number of hidden layers. *l1*, *l2* and *input_dropout_ratio* parameters were used to prevent the most significant risk of DL analysis namely overfitting of the data. The objective of performing *input_dropuout_ratio* parameter is to arbitrarily decline units from the neural network during the training phase. This avoids incorporation of units due to co-adaptation and minimizes overfitting.

4.4.6. Ranking important features

A model-based approach was considered to determine the contribution of a particular 'feature' (predictor) to the model performance. We ranked the importance of the predictive features (in decreasing order) in each of the six AI predictive algorithms by using the variable importance functions *varimp* in h2o and *varImp* in caret R packages. The scripts for the AI analysis has been provided in a Supplemental file-Sample Source Code of AI-ML/DL Models.

4.5. Multivariate regression analysis

We also evaluated the predictive accuracy of DNA methylation changes using a conventional logistic regression approach. Prior to development of a logistic regression model for detection of autism, raw epigenetics data were subjected to PQN normalization, log transformation and auto-scaling. To select the best set of predictor variables utilized in the model and for optimizing all of the model components, stepwise variable selection and Least Absolute Shrinkage and Selection Operator (LASSO) were utilized. Following the 10-fold cross-validation, model performance metrics that included AUC (area under the receiver operating characteristics curve), specificity and sensitivity were generated.

4.6. Heatmap

The differentially methylated CpG sites in the promoter and/or coding regions were subsequently used to generate a heatmap using the R package (v3.2.2) ComplexHeatmap (v1.6.0) module. For the hierarchical clustering of samples, we used ward distance method (Gu, 2015). FDR p-value, methylation status and the AUC were used to distinguish autism cases from controls. The methylation level in the most significantly differentially methylated CpG loci were considered to calculate area under the receiver operating characteristic (AUC-ROC).

Data cleaning, normalization and analysis was performed using R (version 3.2.3) in RStudio (version 0.99.489). The p-value for methylation differences of each CpG loci between the case and control groups calculated as earlier described (Altorok et al., 2014). Filtering criteria for p-values defining significant methylation differences at each CpG locus were based on the Benjamini - Hochberg False Discovery Rate (FDR) correction for multiple testing (Benjamini and Hochberg, 1995) with p-value threshold set at < 0.05 and with AUC being statistically significant > 0.50 with significant 95% confidence intervals. We also used stringent p-value thresholds set at $p < 5 \times 10^{-8}$. To determine the diagnostic accuracy of specific methylation sites, Area under the characteristic reciever operating curves (AUC-ROC) were calculated.

4.7. Gene functional and pathway enrichment

Pathway and functional enrichment analysis of the genes consisting the differentially methylated CpG loci was performed to investigate plausible biological significance of the methylation changes. The differentially methylated genes with FDR p-value < 0.05 were used for a pathway analysis. Only genes with Entrez identifiers were analyzed. Pathway analysis was carried out using the Ingenuity Pathway Analysis (Ingenuity Systems, www.ingenuity.com). Over-represented canonical pathways, molecular processes and biological processes were identified.

Declaration of conflicting interests

The data generated in this study was used for a provisional patent application.

Acknowledgments

We thank the Michigan Department of Health and Human Services for specimens. We also thank Rose Callahan, Beaumont Health, for critical manuscript review and editing.

Funding

This research was not supported by grants from any funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brainres.2019.146457.

References

Association A.P. 1994. Diagnostic and Statistical Manual of Mental Disorders. 4th ed. Washington, DC.

Alakwaa, F.M., Chaudhary, K., Garmire, L.X., 2018. Deep learning accurately predicts estrogen receptor status in breast cancer metabolomics data. J. Proteome Res. 17, 337–347.

Alpay Savasan, Z., Yilmaz, A., Ugur, Z., Aydas, B., Bahado-Singh, R.O., Graham, S.F., 2019. Metabolomic profiling of cerebral palsy brain tissue reveals novel central biomarkers and biochemical pathways associated with the disease: a pilot study. Metabolites 9.

Altorok, N., Tsou, P.S., Coit, P., Khanna, D., Sawalha, A.H., 2014. Genome-wide DNA methylation analysis in dermal fibroblasts from patients with diffuse and limited systemic sclerosis reveals common and subset-specific DNA methylation aberrancies. Ann. Rheum. Dis.

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders, 5th ed. American Psychiatric Association, Washington, DC.

Anitha, A., Nakamura, K., Thanseem, I., Yamada, K., Iwayama, Y., Toyota, T., Matsuzaki, H., Miyachi, T., Yamada, S., Tsujii, M., Tsuchiya, K.J., Matsumoto, K., Iwata, Y., Suzuki, K., Ichikawa, H., Sugiyama, T., Yoshikawa, T., Mori, N., 2012. Brain region-specific altered expression and association of mitochondria-related genes in autism. Mol. Autism. 3, 12.

Armstrong, C., 2008. AAP releases guidelines on identification of children with autism spectrum disorders. Am. Fam. Physician 78, 1301–1305.

Bahado-Singh, R.O., Sonek, J., McKenna, D., Cool, D., Aydas, B., Turkoglu, O., Bjorndahl, T., Mandal, R., Wishart, D., Friedman, P., Graham, S.F., Yilmaz, A., 2018. Artificial Intelligence and amniotic fluid multiomics analysis: the prediction of perinatal outcome in asymptomatic short cervix. Ultrasound Obstet. Gynecol.

Bahado-Singh, R.O., Vishweswaraiah, S., Aydas, B., Mishra, N.K., Guda, C., Radhakrishna, U., 2019a. Deep learning/artificial intelligence and blood-based DNA epigenomic prediction of cerebral palsy. Int. J. Mol. Sci. 20.

Bahado-Singh, R.O., Yilmaz, A., Bisgin, H., Turkoglu, O., Kumar, P., Sherman, E., Mrazik, A., Odibo, A., Graham, S.F., 2019b. Artificial intelligence and the analysis of multi-platform metabolomics data for the detection of intrauterine growth restriction. PLoS One 14, e0214121.

Bengio, Y., 2009. Learning deep architectures for AI. Found. Trends® Mach. Learn. 2, 1–127.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statist. Soc. B 57, 289–300.

Bergman, O., Westberg, L., Nilsson, L.G., Adolfsson, R., Eriksson, E., 2010. Preliminary evidence that polymorphisms in dopamine-related transcription factors LMX1A, LMX1B and PITX3 are associated with schizophrenia. Prog. Neuropsychopharmacol. Biol. Psychiatry 34, 1094–1097.

Betancur, C., 2011. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. Brain Res. 1380, 42–77.

Butler, M.G., Rafi, S.K., Hossain, W., Stephan, D.A., Manzardo, A.M., 2015. Whole exome sequencing in females with autism implicates novel and candidate genes. Int. J. Mol. Sci. 16, 1312–1335.

Chen, Y.A., Lemire, M., Choufani, S., Butcher, D.T., Grafodatskaya, D., Zanke, B.W., Gallinger, S., Hudson, T.J., Weksberg, R., 2013. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 8, 203–209.

Coury, D.L., Ashwood, P., Fasano, A., Fuchs, G., Geraghty, M., Kaul, A., Mawe, G., Patterson, P., Jones, N.E., 2012. Gastrointestinal conditions in children with autism

spectrum disorder: developing a research agenda. Pediatrics 130 (Suppl 2), S160–S168.
Daca-Roszak, P., Pfeifer, A., Zebracka-Gala, J., Rusinek, D., Szybinska, A., Jarzab, B., Witt, M., Zietkiewicz, E., 2015. Impact of SNPs on methylation readouts by Illumina Infinium HumanMethylation450 BeadChip Array: implications for comparative population studies. BMC Genomics 16, 1003.
Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., Donaldson, A., Varley, J., 2010. Randomized, controlled trial of an intervention for toddlers with autism: the early start denver model. Pediatrics 125, e17–e23.
Dipasquale, V., Cutrupi, M.C., Colavita, L., Manti, S., Cuppari, C., Salpietro, C., 2017. Neuroinflammation in autism spectrum disorders: role of high mobility group box 1 protein. Int. J. Mol. Cell. Med. 6, 148–155.
Docherty, L.E., Rezwan, F.I., Poole, R.L., Jagoe, H., Lake, H., Lockett, G.A., Arshad, H., Wilson, D.I., Holloway, J.W., Temple, I.K., Mackay, D.J., 2014. Genome-wide DNA methylation analysis of patients with imprinting disorders identifies differentially methylated regions associated with novel candidate imprinted genes. J. Med. Genet. 51, 229–238.
Govek, E.E., Newey, S.E., Van Aelst, L., 2005. The role of the Rho GTPases in neuronal development. Genes Dev. 19, 1–49.
Grice, D.E., Buxbaum, J.D., 2006. The genetics of autism spectrum disorders. Neuromol. Med. 8, 451–460.
Gu, Z., 2015. ComplexHeatmap: Making Complex Heatmaps. R package version 1 (6). https://github.com/jokergoo/ComplexHeatmap.
Hu, V.W., 2013. From genes to environment: using integrative genomics to build a "systems-level" understanding of autism spectrum disorders. Child Dev. 84, 89–103.
Johnson, C.P., Myers, S.M., American Academy of Pediatrics Council on Children With, D., 2007. Identification and evaluation of children with autism spectrum disorders. Pediatrics 120, 1183–1215.
Joo, J.E., Wong, E.M., Baglietto, L., Jung, C.H., Tsimiklis, H., Park, D.J., Wong, N.C., English, D.R., Hopper, J.L., Severi, G., Giles, G.G., Southey, M.C., 2013. The use of DNA from archival dried blood spots with the Infinium HumanMethylation450 array. BMC Biotechnol. 13, 23.
Kasari, C., Gulsrud, A.C., Wong, C., Kwon, S., Locke, J., 2010. Randomized controlled caregiver mediated joint engagement intervention for toddlers with autism. J. Autism Dev. Disord. 40, 1045–1056.
Kavalali, E.T., Nelson, E.D., Monteggia, L.M., 2011. Role of MeCP2, DNA methylation, and HDACs in regulating synapse function. J. Neurodev. Disord. 3, 250–256.
Krumm, N., O'Roak, B.J., Karakoc, E., Mohajeri, K., Nelson, B., Vives, L., Jacquemont, S., Munson, J., Bernier, R., Eichler, E.E., 2013. Transmission disequilibrium of small CNVs in simplex autism. Am. J. Hum. Genet. 93, 595–606.
Ladd-Acosta, C., Hansen, K.D., Briem, E., Fallin, M.D., Kaufmann, W.E., Feinberg, A.P., 2014. Common DNA methylation alterations in multiple brain regions in autism. Mol. Psychiatry 19, 862–871.
Lee, D.Y., 2015. Roles of mTOR signaling in brain development. Exp. Neurobiol. 24, 177–185.
Leenen, F.A., Muller, C.P., Turner, J.D., 2016. DNA methylation: conducting the orchestra from exposure to phenotype? Clin. Epigenet. 8, 92.
Liu, Y., Aryee, M.J., Padyukov, L., Fallin, M.D., Hesselberg, E., Runarsson, A., Reinius, L., Acevedo, N., Taub, M., Ronninger, M., Shchetynsky, K., Scheynius, A., Kere, J., Alfredsson, L., Klareskog, L., Ekstrom, T.J., Feinberg, A.P., 2013. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat. Biotechnol. 31, 142–147.
Loke, Y.J., Hannan, A.J., Craig, J.M., 2015. The role of epigenetic change in autism spectrum disorders. Front. Neurol. 6, 107.
Lopez-Rangel, E., Lewis, M.E., 2006. Loud and clear evidence for gene silencing by epigenetic mechanisms in autism spectrum and related neurodevelopmental disorders. Clin. Genet. 69, 21–22.
Matelski, L., Van de Water, J., 2016. Risk factors in autism: thinking outside the brain. J. Autoimmun. 67, 1–7.
McCarthy, M., 2014. Autism diagnoses in the US rise by 30%, CDC reports. BMJ 348, g2520.
Mefford, H.C., Batshaw, M.L., Hoffman, E.P., 2012. Genomics, intellectual disability, and autism. N. Engl. J. Med. 366, 733–743.
Mendelsohn, N.J., Schaefer, G.B., 2008. Genetic evaluation of autism. Semin. Pediatr. Neurol. 15, 27–31.
Nagarajan, R.P., Hogart, A.R., Gwye, Y., Martin, M.R., LaSalle, J.M., 2006. Reduced MeCP2 expression is frequent in autism frontal cortex and correlates with aberrant MECP2 promoter methylation. Epigenetics 1, e1–e11.
Nakagawa, Y., Chiba, K., 2016. Involvement of neuroinflammation during brain development in social cognitive deficits in autism spectrum disorder and schizophrenia. J. Pharmacol. Exp. Ther. 358, 504–515.
Narumiya, S., Thumkeo, D., 2018. Rho signaling research: history, current status and future directions. FEBS Lett. 592, 1763–1776.
Nevison, C.D., 2014. A comparison of temporal trends in United States autism prevalence to trends in suspected environmental factors. Environ. Health 13, 73.
Onore, C., Yang, H., Van de Water, J., Ashwood, P., 2017. Dynamic Akt/mTOR signaling in children with autism spectrum disorder. Front. Pediatr. 5, 43.
Pan, H., Chen, L., Dogra, S., Teh, A.L., Tan, J.H., Lim, Y.I., Lim, Y.C., Jin, S., Lee, Y.K., Ng, P.Y., Ong, M.L., Barton, S., Chong, Y.S., Meaney, M.J., Gluckman, P.D., Stunkel, W., Ding, C., Holbrook, J.D., 2012. Measuring the methylome in clinical samples: improved processing of the Infinium Human Methylation450 BeadChip Array. Epigenetics 7, 1173–1187.
Radhakrishna, U., Albayrak, S., Alpay-Savasan, Z., Zeb, A., Turkoglu, O., Sobolewski, P., Bahado-Singh, R.O., 2016. Genome-wide DNA methylation analysis and epigenetic variations associated with congenital aortic valve stenosis (AVS). PLoS One 11, e0154010.
Radhakrishna, U., Vishweswaraiah, S., Veerappa, A.M., Zafra, R., Albayrak, S., Sitharam, P.H., Saiyed, N.M., Mishra, N.K., Guda, C., Bahado-Singh, R., 2018. Newborn blood DNA epigenetic variations and signaling pathway genes associated with Tetralogy of Fallot (TOF). PLoS One 13, e0203893.
Reddy, K.S., 2005. Cytogenetic abnormalities and fragile-X syndrome in autism spectrum disorder. BMC Med. Genet. 6, 3.
Redies, C., Hertel, N., Hubner, C.A., 2012. Cadherins and neuropsychiatric disorders. Brain Res. 1470, 130–144.
Risch, N., Merikangas, K., 1996. The future of genetic studies of complex human diseases. Science 273, 1516–1517.
Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., Moreno-De-Luca, D., Chu, S.H., Moreau, M.P., Gupta, A.R., Thomson, S.A., Mason, C.E., Bilguvar, K., Celestino-Soper, P.B., Choi, M., Crawford, E.L., Davis, L., Wright, N.R., Dhodapkar, R.M., DiCola, M., DiLullo, N.M., Fernandez, T.V., Fielding-Singh, V., Fishman, D.O., Frahm, S., Garagaloyan, R., Goh, G.S., Kammela, S., Klei, L., Lowe, J.K., Lund, S.C., McGrew, A.D., Meyer, K.A., Moffat, W.J., Murdoch, J.D., O'Roak, B.J., Ober, G.T., Pottenger, R.S., Raubeson, M.J., Song, Y., Wang, Q., Yaspan, B.L., Yu, T.W., Yurkiewicz, I.R., Beaudet, A.L., Cantor, R.M., Curland, M., Grice, D.E., Gunel, M., Lifton, R.P., Mane, S.M., Martin, D.M., Shaw, C.A., Sheldon, M., Tischfield, J.A., Walsh, C.A., Morrow, E.M., Ledbetter, D.H., Fombonne, E., Lord, C., Martin, C.L., Brooks, A.I., Sutcliffe, J.S., Cook Jr., E.H., Geschwind, D., Roeder, K., Devlin, B., State, M.W., 2011. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron 70, 863–885.
Schmidhuber, J., 2015. Deep learning in neural networks: an overview. Neural Networks 61, 85–117.
Sealey, L.A., Hughes, B.W., Sriskanda, A.N., Guest, J.R., Gibson, A.D., Johnson-Williams, L., Pace, D.G., Bagasra, O., 2016. Environmental factors in the development of autism spectrum disorders. Environ. Int. 88, 288–298.
Thanseem, I., Nakamura, K., Anitha, A., Suda, S., Yamada, K., Iwayama, Y., Toyota, T., Tsujii, M., Iwata, Y., Suzuki, K., Matsuzaki, H., Iwata, K., Sugiyama, T., Yoshikawa, T., Mori, N., 2011. Association of transcription factor gene LMX1B with autism. PLoS One 6, e23738.
Vahia, V.N., 2013. Diagnostic and statistical manual of mental disorders 5: a quick glance. Indian J. Psychiatry 55, 220–223.
Voineagu, I., Wang, X., Johnston, P., Lowe, J.K., Tian, Y., Horvath, S., Mill, J., Cantor, R.M., Blencowe, B.J., Geschwind, D.H., 2011. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature 474, 380–384.
Wang, L., 2017. Data mining, machine learning and big data analytics. Int. Trans. Electr. Comput. Eng. Syst. 4, 55–61.
Wilhelm-Benartzi, C.S., Koestler, D.C., Karagas, M.R., Flanagan, J.M., Christensen, B.C., Kelsey, K.T., Marsit, C.J., Houseman, E.A., Brown, R., 2013. Review of processing and analysis methods for DNA methylation array data. Br. J. Cancer 109, 1394–1402.
Winden, K.D., Ebrahimi-Fakhari, D., Sahin, M., 2018. Abnormal mTOR Activation in Autism. Ann. Rev. Neurosci. 41, 1–23.
Wong, N., Morley, R., Saffery, R., Craig, J., 2008. Archived Guthrie blood spots as a novel source for quantitative DNA methylation analysis. Biotechniques 45–423424 426, 428 passim.
Xiao, Z., Qiu, T., Ke, X., Xiao, X., Xiao, T., Liang, F., Zou, B., Huang, H., Fang, H., Chu, K., Zhang, J., Liu, Y., 2014. Autism spectrum disorder as early neurodevelopmental disorder: evidence from the brain imaging abnormalities in 2–3 years old toddlers. J. Autism. Dev. Disord. 44, 1633–1640.
Zwaigenbaum, L., Bauman, M.L., Fein, D., Pierce, K., Buie, T., Davis, P.A., Newschaffer, C., Robins, D.L., Wetherby, A., Choueiri, R., Kasari, C., Stone, W.L., Yirmiya, N., Estes, A., Hansen, R.L., McPartland, J.C., Natowicz, M.R., Carter, A., Granpeesheh, D., Mailloux, Z., Smith Roley, S., Wagner, S., 2015. Early screening of autism spectrum disorder: recommendations for practice and research. Pediatrics 136 (Suppl 1), S41–S59.