A Metabolic Profile of Autism Spectrum Disorder from Autism Phenome Project Patient Plasma

R. Burrier[1], A. Smith[1], M. Ross[1], B. Fontaine[1], P. West[1], S. Rogers[2], D. Li[2], D. Amaral[3], and E. Donley[1]
[1]Stemina Biomarker Discovery Inc., 504 S. Rosa Rd., Suite 150, Madison WI 53719
[2]University of California, Davis M.I.N.D. Institute

Overview

PURPOSE: Discovery of metabolic biomarkers for detection of ASD in children.

METHODS: MS-based metabolomic analysis with univariate and multivariate data modelling.

RESULTS: Biomarkers that could properly classify the ASD and TD patients with 79% accuracy.

Introduction

The diagnosis of autism spectrum disorder (ASD) at the earliest age possible is important for initiating optimally effective intervention. Patients can be reliably diagnosed through behavioral testing at about 2 years of age. However, in the United States the average age of diagnosis is around 4 years. Identifying metabolic biomarker signatures of ASD from blood samples offers an opportunity for developing earlier diagnostic tests.

Objectives

- Discover metabolic features in plasma samples that can be used as biomarkers to discriminate children with ASD from typically developing (TD) children.
- Evaluate these biomarkers in an independent set of patient samples.
- Explore potential metabolic subtypes with ASD.
- Confirm the chemical structures of the biomarkers.

Methods

Subject Samples

- Diagnosis of autism using ADOS-G and ADI-R and criteria from the Collaborative Programs of Excellence in Autism
- TD children included if developmental scores were within 2 standard deviations of the mean on all subscales of the MSEL. TD exclusion criteria included mental retardation, pervasive developmental disorder, language impairment or other developmental, neurological, or behavioral problems. TD children were screened and excluded for autism with the Social Communication Questionnaire.
- Non-fasted blood was obtained in ACD tubes and the plasma was stored at -80°C.

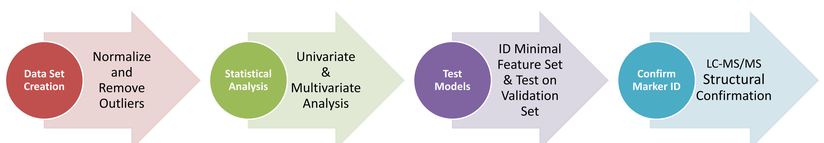
Patient demographics

Demographic		TD	ASD	Overall
Group Size		93	180	273
Sex (male %)		69	83	78
Age (y)	Ave ± S.D.	3.0±0.4	3.1±0.5	3.0±0.5
DQ	Ave ± S.D.	106.4±11.9	62.5±20.8	77.3±27.8

Sample Preparation and Mass Spectrometry

- Small molecules extracted using 8:1 methanol:water solution at -20°C.
- Samples were centrifuged to remove precipitate, evaporated to dryness then solubilized for LC-HRMS analysis.
- Untargeted LC-HRMS (C8 or HILIC chromatography) methods were optimized for metabolome coverage. LC-HRMS was performed using an Agilent G6520 QTOF LC-HRMS system.
- Electrospray ionization (ESI) in both positive and negative ion modes under high-resolution exact mass conditions.
- 15 of the samples did not pass MS quality review and were removed from further analysis.
- GC-MS was performed at the West Coast Metabolomics Center

Data Analysis



- Combine data across LC and GC MS methods
- Partition samples into training and validation sets

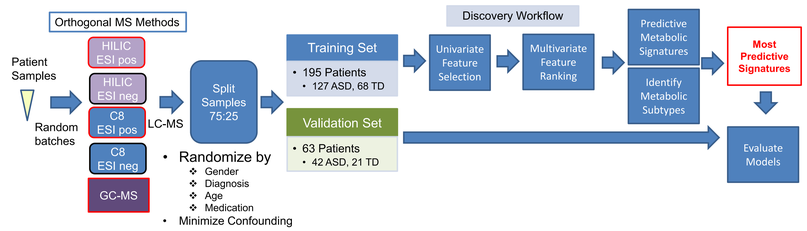
- T-test feature selection
- Heuristic based feature selection
- Supervised methods with nested resampling for VIPs calculation

- VIP feature ranking.
- Select best performing set

- Chemical reference library evaluation
- Confirm LC-MS/MS

Results

Experimental Workflow



Feature Contribution by Analytical Platform

Platform	Features	After QC	P-value <= 0.05
HILIC +	2629	653	113
HILIC -	2364	565	71
C8 +	758	301	42
C8 -	736	235	66
GCMS	378	378	22
Total	6865	2132	314

Methods were optimized so that each of the analytical methods contributed to the overall orthogonal analytical approach. After QC, there were 314 features that were carried forward for classification modeling.

Classification Results for Both Training and Validation Sample Sets

Three computational modeling methods PLS-DA, random forest (RF), and SVM were used to calculate VIP scores and to select the optimal modeling method and feature subset. Random forest yielded the best performing models having good accuracy, excellent sensitivity, and marginal specificity. The best performing model was chosen based on prediction of the validation set.

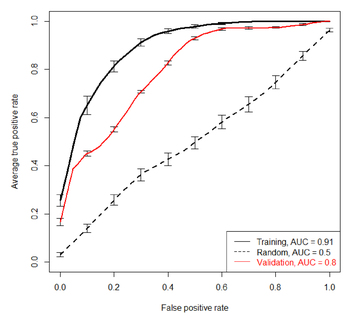
Training Set Results

Model	Feature No.	Accuracy	Sensitivity	Specificity	AUC
RF	120	0.82	0.98	0.52	0.91

Independent Validation Set Results

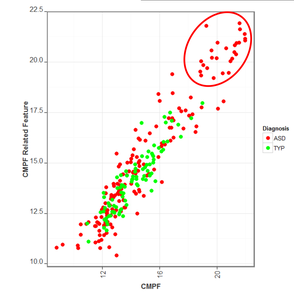
Model	Feature No.	Accuracy	Sensitivity	Specificity	AUC
RF	120	0.79	0.84	0.48	0.80

ROC Analysis for Training and Validation Data Sets

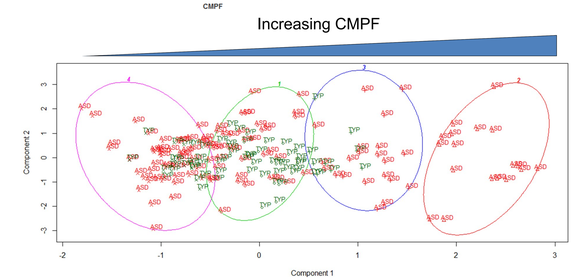


ROC analysis of the top performing model. The training set results (**black, solid**) are based on the average of the prediction from the hold out samples of 5-fold cross validation repeated 5 times. The validation data (**red**) were generated from the prediction of the validation set samples using 5 random seeds. A null model (**black, dotted**) was created by randomizing diagnosis to demonstrate that the ASD vs TD classification results were not obtained by chance.

CMPF Identified as a Biomarker of an ASD Metabolic Subtype



- K-means evaluation of CMPF in the training set and prediction of the clusters in the validation set demonstrate a reproducible signature.
- High positive value (PPV > 0.95) with CMPF



K-Means	Training Set		Validation Set Prediction	
Cluster	ASD	TYP	ASD	TYP
1	34	37	14	7
2	**18**	**0**	**10**	**0**
3	21	10	3	5
4	54	21	15	9

- 3-Carboxy-4-methyl-5-propyl-2-furanpropionic acid (CMPF) is a member of the uremic toxins.
- Simple abundance threshold can distinguish a subpopulation of 14 % of the individuals with ASD while an unsupervised k-means approach can describe 24% of the ASD population in this study.
- CMPF inhibits OAT3 transporters and is associated with neurological changes at high plasma concentrations.

Confirmed Features

Metabolite	LC-MS Method	Fold (ASD/TD)	P-Value	FDR
2-Hydroxy-2-methylbutyric acid	C8 ESIneg	0.84	0.006	0.282
3-Methyl-2-oxovaleric acid	C8 ESIneg	0.87	0.013	0.412
Salicylic acid	C8 ESIneg	0.77	0.002	0.174
Gentisic acid	C8 ESIneg	0.71	0.000	0.036
CMPF related metabolite	C8 ESIneg	8.43	0.046	0.637
DHEA sulfate	C8 ESIneg	2.63	0.001	0.127
Pregnenolone sulfate	C8 ESIneg	1.71	0.000	0.029
LysoPE(22:6)	C8 ESIpos	1.38	0.000	0.028
Glycine	HILIC ESIpos	1.34	0.000	0.069
L-Alanine or Sarcosine	HILIC ESIpos	1.17	0.005	0.255
Proline betaine	HILIC ESIpos	0.72	0.000	0.069

Confirmed metabolites present in the top performing predictive model. Values based on the training set of samples. Feature annotations based on retention time match to a chemical reference standard in the Stemina Metabolite Database or a spectral match based on MS-MS fragmentation compared to a public database.

Conclusions

- We demonstrated that 120 features with differential abundance in ASD vs Typical patients (2-4 years old) can be used to derive classification models that can discriminate ASD from TD individuals with 79% accuracy.
- The metabolites identified contain both known ASD biomarkers as well as some new biomarkers.
- Classes of metabolites include lysophospholipids, organic acids, hormone sulfates, furans, and amino acids.
- CMPF may represent a biomarker associated with a metabolic subtype of ASD.
- Discovery of additional subtypes will require larger patient populations.

Future plans include:

1) Explore the biological relationship of CMPF to ASD.
2) Develop subtypes into a panel of diagnostic tests which may enable personalized treatment.
3) Launch Children's Autism Metabolome Project (CAMP) to discover metabolic subtypes of ASD.