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. Lu 1, D. Amaral 2, 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 optimal 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 subcales 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 any with the Social Communication Questionnaire.
- Non-fasted blood was obtained in ACD tubes and the plasma was stored at -80°C.

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 metabolite coverage. LC-HRMS was performed using an Agilent 6520 QTOF LC-HRMS system.
- Electrospray ionization (ESI) in both positive and negative ion modes under high HRMS was performed using an Agilent G6520 QTOF LC-MS/MS.
- Normalization and removal of outliers using the QuanExpress2012 software from Stemina Biomarker Discovery.
- Statistical analysis was performed using VIP feature ranking.
- Statistical analysis was performed using randomizing by gender, diagnosis, age, and medication.
- Confounding was minimized.

Data Analysis

- Metabolite Database or a spectral match based on MS to confirm metabolites present in the top performing predictive model. Values based on the training set of 42 ASD, 21 TD patients.
- ROC curves were generated from the prediction of the validation set using 5 randomized seeds.

Results

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 marginally specificity. The best performing model was chosen based on the prediction of the validation set.

<table>
<thead>
<tr>
<th>Model</th>
<th>Feature No.</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF</td>
<td>120</td>
<td>0.82</td>
<td>0.98</td>
<td>0.52</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Independent Validation Set Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Feature No.</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF</td>
<td>120</td>
<td>0.70</td>
<td>0.64</td>
<td>0.40</td>
<td>0.80</td>
</tr>
</tbody>
</table>

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 using 5 random seeds. A null model (black, dotted) was created by randomizing to demonstrate that the ASD vs TD classification results were not obtained by chance.

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.