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.

Perinatal Steroid Signature in Autism: A Study in the Danish National Birth Cohort

Jane English

jane.english@ucc.ie

University College Cork https://orcid.org/0000-0001-9930-7050

Sebastian Dohm-Hansen

https://orcid.org/0000-0002-3309-9380

Aisling Noone

University College Cork https://orcid.org/0009-0004-9799-1951

Kirsten Dowling

Eric Wolsztynski

Caitriona Scaife

Anna Golubeva

University College Cork https://orcid.org/0000-0003-0112-6951

Louise Gallagher

Tine Henriksen

https://orcid.org/0000-0001-6933-8294

Ali Khashan

University College Cork

Bodil Bech

Article

Keywords: Autism, Danish National Birth Cohort (DNBC), steroids, allopregnanolone, pregnenolone, extracellular matrix (ECM)

Posted Date: September 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10315219/v1

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

Additional Declarations: The authors have declared there is **NO** conflict of interest to disclose

Abstract

Autism has prenatal origins, yet diagnosis typically occurs years after behavioural signs emerge. Umbilical cord blood (UCB) provides a unique window into the perinatal molecular environment, but large-scale studies examining molecular signatures associated with subsequent autism diagnosis remain limited. Here, we performed metabolomic and proteomic profiling of UCB in a nested case-control study within the Danish National Birth Cohort. Among 70,292 pregnancies, we identified 465 mother–child dyads with childhood autism (ICD-10 F84.0) and 465 neurotypical (NT) controls. UCB samples were available from 398 autistic and 394 NT children, which were divided into two largely orthogonal participant sets to enable initial profiling and subsequent confirmation of key pathway findings. Participant set 1 (258 autism, 263 NT controls) underwent multi-modal omics profiling, including untargeted metabolomics, a targeted metabolite panel, and discovery proteomics, while participant set 2 (140 autism, 131 NT controls) underwent discovery proteomics and targeted steroid metabolomics. Metabolomics revealed differences in fatty acid, phospholipid, and steroid profiles between autistic and NT children. A reduced steroid signature emerged in participant set 1, characterized by a significantly lower aggregate steroid score in autistic children compared with NT children ($P = 0.00517$). In participant set 2, targeted steroid analysis confirmed reduced concentrations of the neuroactive steroid allopregnanolone ($P = 0.0279$), and pregnenolone ($P = 0.0300$), with modest discrimination between autistic and NT children (AUROC = 0.64, 95% CI 0.59–0.67). Proteomic analyses independently identified convergent alterations in complement signalling and extracellular matrix signalling across both participant subsets. These findings identify a reproducible perinatal steroid signature associated with later autism diagnosis, and implicate disrupted steroidogenesis alongside immune and extracellular matrix signalling in autism biology. These molecular signatures may support prognostic biomarker development and earlier identification of children with autism.

Introduction

Childhood autism, a neurotype of the autism spectrum condition, is characterized by the presence of differences in social interaction and communication, and restricted, or repetitive behavior, before the age of three years [1]. Routine neonatal screening for autism could substantially reduce time to diagnosis and facilitate early intervention for children with high support needs. Although autism may be reliably diagnosed as early as two years [2], the actual age of diagnosis is often considerably later, with marked variation between countries [3, 4]. Recent data further highlights that girls are disproportionately diagnosed later than boys, with diagnostic peaks occurring in adolescence rather than in childhood [5]. Accordingly, many autistic females are not identified until adolescence [6], underscoring the need for early, equitable detection strategies to enable timely support across all subgroups. Delayed diagnosis translates into missed opportunities, as interventions as young as six months old can improve social communication, language, behavior and cognition [7–9], underscoring an urgent need for prognostic biomarkers and clinical decision support systems for the early identification of children with autism.

The Danish National Birth Cohort (DNBC) [10] was established to investigate the causal link between exposures in early life and disease outcomes. The cohort includes prospectively collected prenatal exposure data which can be linked to Danish national health registries to determine neurodevelopmental outcomes. To determine whether biological signatures present at birth are associated with a childhood autism diagnosis, we analyzed maternal health record data and UCB samples from the DNBC using a multi-modal omics approach. We hypothesized that this strategy may help identify convergent perinatal molecular pathways implicated in autism biology and reveal new avenues for supportive early intervention.

Materials and methods

Study design and participants

The DNBC collected data on a large sample of babies born in Denmark between 1997 and 2003. We restricted our analysis to live-born children whose mothers participated in all three telephone interviews up to six months after birth, where information on prenatal health, wellbeing and pregnancy complications had been collected (**Supplementary materials and methods in Supplement 1**). The data were matched with child health outcomes, including childhood autism, through the Danish National Patient Register. In total, 835 children in the DNBC cohort were diagnosed with childhood autism (ICD-10 F84.0) [11]. Of these, we identified 465 mother-child pairs with UCB samples and data available from interviews 1–3. We then randomly selected 465 mother-child pairs as a neurotypical (NT) comparison group, matched 1:1 on date of birth (DOB) and sex. Within the autism group, we identified (through the Danish National Patient Registry) 12 children additionally presenting with Intellectual Disability (ID) (ICD-10 F70.0-F79.0), 119 children with ADHD (ICD-10 F90.0), and 117 children with co-occurring psychiatric conditions. After excluding hemolyzed UCB samples, 398 autism mother-child pairs and 394 NT mother-child pairs remained (**Supplementary Fig. 1 in Supplement 1**). These samples were split into two largely orthogonal sets: Participant dataset 1 was profiled with untargeted and targeted metabolomics and discovery proteomics (four -omics modalities); Participant dataset 2 was profiled with discovery proteomics and a targeted metabolomic assay for steroid compounds (**Supplementary Figs. 1–2 in Supplement 1**). The datasets roughly preserved autism:NT and male:female ratios (**Supplementary Table 1 in Supplement 1**). While samples in dataset 2 were originally stored at -80°C (**Supplementary Table 2 in Supplement 1**), initial piloting revealed no differences secondary to storage temperature (data not shown).

Ethics

Participants of the DNBC were enrolled by informed consent, as detailed on the DNBC website. This study was approved by the Danish Data Protection Agency under the Aarhus University comment agreement (j. number 2015-57-0002) and Aarhus University (j. number 2016-051-000001, sequential

number 1673). The project was granted ethical approval on 12/08/2019 by the Central Denmark Region ethical committee (number: 1-10-72-79-19). The study and its protocol were not preregistered.

Pregnancy and maternal interview data

Traits that have a prior association with autism were extracted from the maternal pregnancy records and interview data. These subject characteristics included maternal age, paternal age [12, 13], history of mental disorder in mother [14, 15], preterm birth (< 37 gestational weeks) [16, 17], preeclampsia [18, 19], maternal BMI [20, 21], fever during pregnancy [22, 23], smoking [24] and alcohol consumption [25] during pregnancy, birthweight [26], and sex of the child [27] (Table 1). We additionally extracted and engineered (**Supplementary materials and methods in Supplement 1**) variables pertaining to environmental exposures (e.g. medication, diet, supplements, stressors), burden of disease, history of miscarriages, and a composite measure of socio-occupational status [28]. These (composite) variables were used for statistical analyses and machine learning.

Table 1
Subject characteristics

Subject Variable	Autism (N = 398)	NT (N = 394)	*p*-value	Participant Set 1 (N = 521)	Participant Set 2 (N = 271)	*p*-value
Indication	-	-	-	Autism' = 258 (50%) 'NT' = 263 (50%)	Autism' = 140 (52%) 'NT' = 131 (48%)	6.19 × 10 − 1
Maternal age	Mean: 29.7 SD: 4.70 N: 398	Mean: 30.7 SD: 4.10 N: 394	**1.84 × 10 − 3**	Mean: 30.2 SD: 4.36 N: 521	Mean: 30.2 SD: 4.59 N: 271	9.08 × 10 − 1
Paternal age (years)	Mean: 31.8 SD: 5.41 N: 395	Mean: 32.6 SD: 4.94 N: 385	**3.85 × 10 − 2**	Mean: 32.2 SD: 5.33 N: 518	Mean: 32.3 SD: 4.93 N: 262	7.28 × 10 − 1
Maternal BMI	Mean: 25.6 SD: 4.84 N: 356	Mean: 24.8 SD: 4.18 N: 365	**1.66 × 10 − 2**	Mean: 25. SD: 4.58 N: 479	Mean: 25.7 SD: 4.41 N: 242	5.63 × 10 − 2
Maternal History of Miscarriage (No/Yes)	No' = 286 (75%) 'Yes' = 94 (25%)	No' = 299 (79%) 'Yes' = 78 (21%)	2.14 × 10 − 1	No' = 382 (76%) 'Yes' = 121 (24%)	No' = 203 (80%) 'Yes' = 51 (20%)	2.54 × 10 − 1
Burdened by bad mood	Mean: 2.98 SD: 2.31 N: 398	Mean: 2.38 SD: 2.04 N: 394	**3.67 × 10 − 4**	Mean: 2.81 SD: 2.22 N: 521	Mean: 2.41 SD: 2.14 N: 271	**2.35 × 10 − 2**
Burdened by disease	Mean: 2.20 SD: 1.86 N: 398	Mean: 1.75 SD: 1.49 N: 394	**1.71 × 10 − 4**	Mean: 1.93 SD: 1.65 N: 521	Mean: 2.05 SD: 1.80 N: 271	3.50 × 10 − 1
Burdened by life	Mean: 1.91 SD: 1.60 N: 398	Mean: 1.42 SD: 1.33 N: 394	**7.82 × 10 − 6**	Mean: 1.69 SD: 1.50 N: 521	Mean: 1.61 SD: 1.47 N: 271	5.13 × 10 − 1
Fever (No/Yes)	No' = 263 (67%)	No' = 272 (70%) 'Yes' =	4.09 × 10 − 1	No' = 358 (69%) 'Yes' = 158 (31%)	No' = 177 (67%) 'Yes' = 89 (33%)	4.67 × 10 − 1

Footnote: Neurotypical controls (NT) were matched on infant sex and date of birth. Numeric values are represented by Mean (M), Standard Deviation (SD), and Number (N) are indicated for each variable. Chi-Square tests were used to assess differences in the categorical variables (No/Yes). Welch's T-tests were performed to assess differences in continuous numeric values ($p < 0.05$). GLM (NB) tests were used to assess differences in count-based variables (burdens) during pregnancy. *p*-values in bold are significant at the $p < .05$ level. Note that missing values have been omitted case-wise in each statistical comparison; thus, numbers may not add up to the totals indicated in the column headers. The “Indication” variable refers to whether the subject is Autism or NT. Socio-occupational status (1 = high SES; 2 = medium SES; 3 = low SES).

Subject Variable	Autism (N = 398)	NT (N = 394)	*p*-value	Participant Set 1 (N = 521)	Participant Set 2 (N = 271)	*p*-value
	'Yes' = 130 (33%)	117 (30%)				
Fish oil (No/Yes)	No' = 365 (95%) 'Yes' = 21 (5%)	No' = 364 (93%) 'Yes' = 26 (7%)	5.72 × 10 − 1	No' = 488 (95%) 'Yes' = 25 (5%)	No' = 241 (92%) 'Yes' = 22 (8%)	7.65 × 10 − 2
Iron pills (No/Yes)	No' = 84 (22%) 'Yes' = 302 (78%)	No' = 64 (16%) 'Yes' = 326 (84%)	7.09 × 10 − 2	No' = 105 (20%) 'Yes' = 408 (80%)	No' = 43 (16%) 'Yes' = 220 (84%)	1.99 × 10 − 1
Medication (No/Yes)	No' = 298 (77%) 'Yes' = 88 (23%)	No' = 313 (81%) 'Yes' = 75 (19%)	2.74 × 10 − 1	No' = 409 (80%) 'Yes' = 104 (20%)	No' = 202 (77%) 'Yes' = 59 (23%)	5.10 × 10 − 1
Maternal History of Mental Health Disorder (No/Yes)	No' = 326 (86%) 'Yes' = 54 (14%)	No' = 362 (96%) 'Yes' = 15 (4%)	**1.90 × 10 − 6**	No' = 454 (90%) 'Yes' = 49 (10%)	No' = 234 (92%) 'Yes' = 20 (8%)	4.78 × 10 − 1
Smoked (No/Yes)	No' = 266 (67%) 'Yes' = 129 (33%)	No' = 309 (79%) 'Yes' = 82 (21%)	**2.99 × 10 − 4**	No' = 382 (74%) 'Yes' = 136 (26%)	No' = 193 (72%) 'Yes' = 75 (28%)	6.64 × 10 − 1
Vitamins and folic (No/Yes)	No' = 21 (5%) 'Yes' = 365 (95%)	No' = 26 (7%) 'Yes' = 364 (93%)	5.72 × 10 − 1	No' = 36 (7%) 'Yes' = 477 (93%)	No' = 11 (4%) 'Yes' = 252 (96%)	1.59 × 10 − 1
Socio-occupational status	1' = 176 (47%) '2' = 142 (38%) '3' = 58 (15%)	1' = 212 (57%) '2' = 133 (35%) '3' = 30 (8%)	**1.89 × 10 − 3**	1' = 254 (51%) '2' = 190 (38%) '3' = 56 (11%)	1' = 134 (53%) '2' = 85 (34%) '3' = 32 (13%)	5.12 × 10 − 1
Gestation days	Mean: 280. SD: 11.9 N: 397	Mean: 280. SD: 10.6 N: 393	8.69 × 10 − 1	Mean: 280. SD: 11.0 N: 519	Mean: 280. SD: 11.8 N: 271	9.75 × 10 − 1

Footnote: Neurotypical controls (NT) were matched on infant sex and date of birth. Numeric values are represented by Mean (M), Standard Deviation (SD), and Number (N) are indicated for each variable. Chi-Square tests were used to assess differences in the categorical variables (No/Yes). Welch's T-tests were performed to assess differences in continuous numeric values ($p < 0.05$). GLM (NB) tests were used to assess differences in count-based variables (burdens) during pregnancy. *p*-values in bold are significant at the $p < .05$ level. Note that missing values have been omitted case-wise in each statistical comparison; thus, numbers may not add up to the totals indicated in the column headers. The “Indication” variable refers to whether the subject is Autism or NT. Socio-occupational status (1 = high SES; 2 = medium SES; 3 = low SES).

Subject Variable	Autism (N = 398)	NT (N = 394)	*p*-value	Participant Set 1 (N = 521)	Participant Set 2 (N = 271)	*p*-value
Birth weight (g): Male	Mean (M): 3600. SD: 558. N: 314	Mean (M): 3678. SD: 576. N: 310	8.72 × 10 − 2	Mean (M): 3632. SD: 562. N: 409	Mean (M): 3653. SD: 580. N: 215	6.69 × 10 − 1
Birth weight (g): Female	Mean (F): 3473. SD: 501. N: 83	Mean (F): 3631. SD: 493. N: 82	**4.31 × 10 − 2**	Mean (F): 3525. SD: 506. N: 109	Mean (F): 3603. SD: 494. N: 56	3.42 × 10 − 1
Preeclampsia (PE)	No' = 387 (97%) 'Yes' = 11 (3%)	No' = 387 (98%) 'Yes' = 7 (2%)	4.88 × 10 − 1	No' = 508 (98%) 'Yes' = 13 (2%)	No' = 266 (98%) 'Yes' = 5 (2%)	7.40 × 10 − 1
Preterm birth (PTB)	No' = 377 (95%) 'Yes' = 20 (5%)	No' = 378 (96%) 'Yes' = 15 (4%)	5.09 × 10 − 1	No' = 498 (96%) 'Yes' = 21 (4%)	No' = 257 (95%) 'Yes' = 14 (5%)	5.86 × 10 − 1
Infant sex (Male/Female)	M' = 315 (79%) 'F' = 83 (21%)	M' = 311 (79%) 'F' = 83 (21%)	1.00	F' = 110 (21%) 'M' = 411 (79%)	F' = 56 (21%) 'M' = 215 (79%)	9.56 × 10 − 1

Footnote: Neurotypical controls (NT) were matched on infant sex and date of birth. Numeric values are represented by Mean (M), Standard Deviation (SD), and Number (N) are indicated for each variable. Chi-Square tests were used to assess differences in the categorical variables (No/Yes). Welch's T-tests were performed to assess differences in continuous numeric values ($p < 0.05$). GLM (NB) tests were used to assess differences in count-based variables (burdens) during pregnancy. *p*-values in bold are significant at the $p < .05$ level. Note that missing values have been omitted case-wise in each statistical comparison; thus, numbers may not add up to the totals indicated in the column headers. The “Indication” variable refers to whether the subject is Autism or NT. Socio-occupational status (1 = high SES; 2 = medium SES; 3 = low SES).

Sample preparation and mass spectrometry analysis for discovery and targeted -omics

For a full multi-modal omics methodological description, please refer to Supplementary materials and methods in Supplement 1.

Statistical Analysis

All analyses in this study were completed using R (version 4.4.0), except for initial -omics data preparation, which was performed in Python with “DEPy” library [29]. For a full list of loaded packages, see **Supplementary Table 4 in Supplement 1**. Univariable Differential Expression Analyses (DEA) of

metabolomics and proteomics data were carried out according to Dohm-Hansen et al. (2026). In short, data were normalized with variance stabilizing normalization (without imputation of missing values), and empirical sample quality weights were estimated using the “arrayWeights” in limma to account for heterogeneous noise. DEA was performed robustly to account for hyper-/hypo-variable features. Maternal and pregnancy-related covariates that were found to have significant effects on -omics modalities were included. The following design matrix was applied independently to each feature: indication (autism:NT), sex, maternal age, preeclampsia, and preterm, where “indication” and “sex” are categorical covariates, and “maternal age” is a numerical encoding chronological age in years, and “preeclampsia” and “preterm birth” are binary encoding their presence/absence (< 37 gestational weeks) **(**Table 1). Features were considered nominally significant at $p < 0.05$, and adjustment for Benjamini-Hochberg-based false discovery rate (FDR) of < 0.05 was applied and reported in tables and figures.

To calculate untargeted steroid aggregate scores, feature-wise standardization was performed followed by aggregation at the subject level using arithmetic mean. In the DNBC cohort, logistic regression was used to predict autism:NT status based on subject aggregate (composite) steroid score in participant set 1, while adjusting for covariates, as per DEA. We hypothesized a directional decrease in the aggregated steroid in the BASELINE Birth Cohort Study, based on prior untargeted metabolomics data from Noone et al. (2026); accordingly, we used a one-sided Welch’s *t*-test. For the targeted steroid data analyses in participant set 2, logistic regression models were fitted to standardized (i.e. z-scored) feature abundance values to predict autism:NT status, while adjusting for sex and batch effects.

Pathway Analysis and Chemical class Enrichment

To explore the broad biological processes altered in UCB plasma from children with autism, all nominally differentially expressed features ($p < 0.05$) from discovery metabolomics and proteomics were analyzed in MetaboAnalyst 6.0 and STRING (version 12.0), respectively.

Supervised machine learning

We followed the TRIPOD checklist for prediction model development (**Supplement 4**) [30]. Tidymodels in R was used to train gradient-boosted decision tree (XGBoost) classifiers predicting offspring autism:NT status based on unimputed, log2-transformed -omics features and sex and/or DNBC interview data. All quantitated features were used for modelling, except in the case of complement proteins, which were chosen *a priori* based on published literature [31]. Models were trained through nested 5x5 cross-validation (CV), stratified by indication to preserve autism:NT ratios across folds. Unbiased performance was estimated using 5 outer folds, while hyperparameters were tuned on 5 inner folds. A Latin hypercube grid with 75 hyperparameter combinations was used for tuning tree depth, learning rate, minimum loss reduction, fraction of features randomly sampled per split, fraction of training data sampled per tree, and L1 and L2 regularization parameters. Binary cross-entropy was used as the training objective. Each inner CV fold was evaluated based on AUROC (Area Under the Receiver Operating Characteristics curve) and used to select the best parameters for outer CV fold fitting and

prediction. Autism-NT classifications were based on default class probability argmax. SHAP values were computed with “SHAPforxgboost” after fitting a model to all training data.

Results

Subject characteristics

We characterized the mother-child dyads based on DNBC interview and pregnancy data (Table 1). Mothers of children with autism tended to report a higher burden of disease, mental and personal life stressors, as well as previous mental health disorders, a history of smoking, and have higher BMI (all $p < 0.05$). Lower socio-occupational status was more frequent among the same mothers ($p < 0.05$). Maternal and paternal age were 1 year younger on average among autism children compared to NT children. No differences were identified for maternal supplement or medication use, fever during pregnancy, history of miscarriage, gestational length, or preeclampsia (all $p > 0.05$). The ratio of autism:NT was identical across sexes (Table 1). Female, but not male, offspring were found to have lower birthweights on average ($p < 0.05$). Finally, there were no meaningful differences between participant datasets 1 and 2 among the same variables, bar a marginal difference in burden of bad mood ($p = 0.03$) **(**Table 1**).**

Differential Expression Analysis

Proteomics

Using discovery proteomics, DEA identified 65 and 46 proteins as nominally differentially expressed ($p < 0.05$) in participant set 1 and participant set 2, respectively (**Supplementary Tables 1–2 in Supplement 2**). Four proteins (MINPP1, MRC2, SERPINB1, C4A) remained significant in participant set 2 following FDR correction (Fig. 3). MINPP1 and MRC2 were increased in autism in comparison to the NT group, while SERPINB1 and C4A were decreased in autism (Fig. 1A). MINPP1 maps to inositol phosphate metabolism. MRC2 is involved in ECM remodeling via its role in endocytosis. SERPINB1 and C4A map to inflammatory and complement pathways which play essential roles in regulating innate immune responses. As participant set 1 and participant set 2 were independent of each other, we explored nominally differentially expressed proteins that were common to both. Three ECM proteins were found to be upregulated (SPARC, TIMP1) or downregulated (CFL1) in autism compared to the NT group across datasets (Fig. 1B). Similarly, at the pathway level, both participant sets revealed complement signaling, and ECM organization among top pathways altered in autism (Fig. 1C). Finally, while DEA analyses were adjusted for sex, the findings were largely driven by effects in males, as evidenced by sex-stratified analyses (**Supplementary Fig. 3A, B in Supplement 1**) **(Supplementary Tables 1–4 in Supplement 3).**

Metabolomics

Using discovery metabolomics, DEA identified 30 features in negative ion mode (-‘ve mode), 123 features in positive ion mode (+’ve mode), and 13 features in the targeted metabolite panel as nominally differential expressed ($p < 0.05$) (Fig. 2A) (**Supplementary Tables 3–5 in Supplement 2**). After FDR

adjustment, four lipids remained significant: Docosapentaenoic acid (FDR = 0.030), Phosphatidylcholine (PC) (42:1) (FDR = 0.042), lysoPC (20:3) (FDR = 0.042) and lysoPC (18:2) (FDR = 0.042) (Fig. 2B). As with the proteomics results, these findings were largely driven by effects in males, as evidenced by sex-stratified analyses (**Supplementary Fig. 3A, B in Supplement 1**) (**Supplementary Tables 5–10 in Supplement 3**). Chemical class enrichment revealed six significantly enriched classes: fatty acyls (FDR < 0.001), carboxylic acids (FDR < 0.001), indoles (FDR = 0.014), steroids (FDR = 0.023), prenol lipids (FDR = 0.023) and organooxygen compounds (FDR = 0.032) (Fig. 2C).

Examining the direction of change across compound classes revealed that steroid features as a group tended to have reduced fold changes in UCB in autism in comparison to NT infants (Fig. 3A). This steroid signature prompted the construct of a standardized aggregated score from all quantified compounds belonging to the steroid chemical class. Logistic regression analysis revealed that the aggregated steroid score was associated with a lower autism odds ratio (β = -0.752, p = 0.00517, OR = 0.471, 95% CI [0.275, 0.793]) (Fig. 3B). These results were broadly consistent with our observations in a previously published metabolomics study in a smaller, demographically independent pregnancy cohort called BASELINE [32]. In that study, we observed nominally reduced expression of 8 steroid compounds in UCB from children with autism in comparison to NT participants. We therefore sought to independently replicate the reduced aggregated steroid profile in the BASELINE cohort. The aggregated steroid score was directionally consistent with our DNBC findings (Fig. 3C). In the Baseline cohort, the autism group showed a numeric reduction in the steroid composite score relative to NT, with a Welch's t = -1.55 (p = 0.0650, one-tailed) (Fig. 3D). To further characterize steroid changes identified through untargeted metabolomics profiling, ten steroid compounds were simultaneously quantified in UCB in DNBC participant set 2, by targeted MRM mass spectrometry. Analysis by univariate logistic regression analysis confirmed reduced Allopregnanolone (β = -0.360, p = 0.0279) and Pregnenolone (β = -0.359, p = 0.0300) in UCB from autistic infants in comparison to NT infants (**Supplementary Table 5 in Supplement 1**).

Supervised machine learning

To capture more complex (non-linear) relationships within our datasets, we trained gradient-boosted, tree-based classifiers using all multi-omic features in participant set 1, in addition to DNBC interview and pregnancy variables (i.e. those in Table 1) but ultimately, no model performed well (**Supplementary Table 6 in Supplement 1**). Given the tentative support for a differential steroid profile in participant set 1, we additionally trained a model using targeted steroid metabolomics data in participant set 2. Using all ten assayed steroids and sex as features, the final model reached an average outer-fold AUROC of 0.64 (95% CI [0.58, 0.70]) (Fig. 4A), with equivalent sensitivity and specificity (**Supplementary Table 6 in Supplement 1**). Adding maternal interview and pregnancy data to the targeted steroid model did not improve performance. The most predictive feature was cortisol, followed closely by allopregnanolone and DHEAS (Fig. 4B). Interestingly, all three features showed evidence of non-linear relationships between feature abundance and predictive strength (**Supplementary Fig. 4A in Supplement 1**), which we further confirmed using Generalized Additive Models (**Supplementary Fig. 4B in Supplement 1**).

Based on the observed enrichment of complement and coagulation-related processes among nominal proteomic DEA features in participant set 1 (Fig. 1C) as well as the findings from recent literature [31], we additionally trained a classifier using an *a priori* selection of complement/coagulation-associated proteins and sex in participant set 2. Although the average outer-fold AUROC was comparable to the targeted steroid-based model (mean = 0.63, 95% CI [0.51, 0.74]) (Fig. 4C), specificity was poor (mean = 0.43, 95% CI [0.35, 0.51]) (**Supplementary Table 6 in Supplement 1**). The top three predictive features were C4A, SERPIND1, and SERPINA4 (Fig. 4D), with less clear relationships between feature importance and abundance (**Supplementary Fig. 4C in Supplement 1**).

Discussion

We employed a nested case-control study design within the DNBC, and a multi-modal omics analysis to investigate if molecular profiles in UCB are associated with a subsequent diagnosis of childhood autism. Using discovery metabolomics, we identified a broad-spectrum reduction in the UCB steroid profile of children with autism, with robust group-level differences in comparison to NT children. A targeted steroid assay confirmed reduced levels of the neuroactive steroids, allopregnanolone and pregnenolone in an orthogonal participant subset, although classification performance of these targeted steroids for autism was modest (AUROC = 0.64). Proteomics further revealed differences in complement signaling and ECM signaling pathways in infants with autism. The concurrent changes of steroids, complement immune activation, and ECM integrity in UCB in autism, points to a placenta-brain axis whereby prenatal steroid insufficiency, as proposed in recent neuroplacentology frameworks [33], drives broader alterations in immune responses and ECM signaling pathways, both critical modulators of neuronal plasticity.

The broad-spectrum reduction in the composite steroid profile observed in this cohort aligns with a growing body of evidence indicating that perinatal molecular profiling can serve as a window into fetal neurobiology [32, 34, 35]. Mechanistically, the placenta functions as a fetal endocrine organ, secreting neuroactive steroids that circulate at high levels during gestation and directly shape brain development [33]. Among these, allopregnanolone modulates GABA-A receptors, and its insufficiency has been shown to disrupt cerebellar white matter, triggering neuroinflammatory and microglial dysfunction, and contribute to autistic-like behaviors in a sex-linked manner [36, 37]. Our findings extend this model to humans, demonstrating that a broad-spectrum reduction in circulating UCB steroids, including the GABAergic neurosteroids allopregnanolone and pregnenolone, in addition to the adrenal androgens dehydroepiandrosterone (DHEA) and androstenedione [32], is detectable at birth. Notably, while others have reported elevated prenatal steroids in autism (amniotic fluid) [38, 39], our findings of reduced steroids in UCB may reflect a dynamic trajectory, from high in mid-gestation to low at birth, highlighting the complexity of prenatal steroidogenic regulation, findings that warrant further investigation. In summary, these systemic changes likely operate in concert with other protein pathways identified by UCB in autism [32, 34, 35], such as immune signaling to shape the final architecture of neural circuits and subsequent neurodevelopmental trajectories.

The present study has both strengths and limitations. To our knowledge, this is the first study to combine maternal health record data, cord blood multi-modal omics data, and machine learning for predicting the likelihood of an autism diagnosis. Our supervised machine learning approach relies on gold standard, unbiased performance estimation through nested CV, whereby hyperparameter tuning and training processes are separated. Data leakage was avoided by filtering missing values, avoiding empirical transformations, and most importantly, not selecting features based on univariate results from the same -omics datasets, thereby limiting overly optimistic performance estimates. Herein, we used all quantitated features identified in our multi-modal discovery -omics data or selected features based on literature and orthogonal datasets. Furthermore, our DEA statistical approach is more robust than traditional methods, such as univariate *t*-tests or ANOVAs [29]. Another key strength is our focus on a homogeneous clinical group, participants with childhood autism (ICD-10 F84.0), rather than a broader autism spectrum diagnosis, as this approach aimed to reduce phenotypic heterogeneity and enhance the detection of biological signals. An important limitation is the gender imbalance of our dataset, a challenge inherent to most studies in autism research as boys are diagnosed more often in childhood. Although not powered to examine sex-specific differences, we found preliminary evidence of divergent signals between sexes, in line with Afonin et al. (2026), who reported plasma protein expression correlating more strongly with externalizing mental health symptoms in males, and with internalizing symptoms in females, consistent with known behavioral sex differences. Moreover, a recent landmark Swedish birth cohort study (n = 2,756,779) found that the male-to-female ratio in autism diagnoses narrows substantially with age and occurs at comparable rates across sexes by adulthood [5]. Therefore, future sex-balanced studies may be achievable by broadening the inclusion criteria beyond childhood autism, to capture females diagnosed during and after adolescence, and improving the generalizability of biomarker discovery efforts to females.

To conclude, our study suggests that altered perinatal steroidogenic regulation is implicated in the neurobiology of autism. These findings point to a placenta-brain axis linking altered steroidogenic regulation to molecular pathways that modulate neuronal plasticity. While further replication and prospective validation are needed, this work lays a foundation for future biomarker research for supportive early intervention.

Declarations

Competing interests

The authors declare no competing interests.

Supplementary information

Supplementary information is available at MP's website.

Author contributions

K.D. and A.N. undertook sample preparation, metabolomics and proteomics analysis of UCB, analyzed the data, prepared figures, and co-wrote the manuscript. S.D-H. undertook ML analysis, statistical analysis, -omics statistical analysis, prepared figures, and co-wrote the manuscript. B.H.B was involved in ethical approval, data management, study design, interpretation of results, and co-wrote the manuscript. A. K., L.G., and T.B.H were involved in study design, interpretation of results, and co-wrote the manuscript. J.E conceived the study, was involved in study design, data management and analysis, interpretation of results, and co-wrote the manuscript. All authors critically reviewed the manuscript and approved the final version for submission.

Acknowledgments

The DNBC was established with a significant grant from the Danish National Research Foundation. Additional support was obtained from the Danish Regional Committees, the Pharmacy Foundation, the Egmont Foundation, the March of Dimes Birth Defects Foundation, the Health Foundation and other minor grants. The DNBC Biobank has been supported by the Novo Nordisk Foundation and the Lundbeck Foundation. This research program was funded by the Irish Health Research Board through an Emerging Investigator Award to Dr Jane English (HRB EIA-2017-023). This manuscript follows the AIMS-2-TRIALS language guidelines for respectful autism research communication [40], consistent with published consensus on identity-first language preferences [41, 42] and neurodiversity-affirming frameworks [42], as well as community guidance from autistic-led organizations [43].

Data availability statement

Data and code are not publicly available because they contain sensitive patient information. Requests for access can be made directly to the DNBC at dnbc-research@ssi.dk.

References

1. World Health Organization. International statistical classification of diseases and related health problems: tenth revision. 2015.
2. Steiner AM, Goldsmith TR, Snow AV, Chawarska K. Practitioner’s Guide to Assessment of Autism Spectrum Disorders in Infants and Toddlers. J Autism Dev Disord. 2012;42:1183–1196.
3. Magán-Maganto M, Bejarano-Martín Á, Fernández-Alvarez C, Narzisi A, García-Primo P, Kawa R, et al. Early Detection and Intervention of ASD: A European Overview. Brain Sciences. 2017;7:159.
4. Malik-Soni N, Shaker A, Luck H, Mullin AE, Wiley RE, Lewis MES, et al. Tackling healthcare access barriers for individuals with autism from diagnosis to adulthood. Pediatr Res. 2022;91:1028–1035.
5. Fyfe C, Winell H, Dougherty J, Gutmann DH, Kolevzon A, Marrus N, et al. Time trends in the male to female ratio for autism incidence: population based, prospectively collected, birth cohort study.

2026. 4 February 2026. https://doi.org/10.1136/bmj-2025-084164.

6. Burrows CA, Sung S, Zheng S, Young GS, Charman T, Klaiman C, et al. Sex-Related Measurement Bias in Autism Spectrum Disorder Symptoms in the Baby Siblings Research Consortium. JAMA Netw Open. 2025;8:e2525887.
7. Estes A, Munson J, Rogers SJ, Greenson J, Winter J, Dawson G. Long-Term Outcomes of Early Intervention in 6-Year-Old Children With Autism Spectrum Disorder. Journal of the American Academy of Child & Adolescent Psychiatry. 2015;54:580–587.
8. Fuller EA, Kaiser AP. The Effects of Early Intervention on Social Communication Outcomes for Children with Autism Spectrum Disorder: A Meta-analysis. J Autism Dev Disord. 2020;50:1683–1700.
9. Mazurek MO, Handen BL, Wodka EL, Nowinski L, Butter E, Engelhardt CR. Age at First Autism Spectrum Disorder Diagnosis: The Role of Birth Cohort, Demographic Factors, and Clinical Features. Journal of Developmental & Behavioral Pediatrics. 2014;35:561.
10. Olsen J, Melbye M, Olsen SF, Sørensen TIA, Aaby P, Nybo Andersen A-M, et al. The Danish National Birth Cohort - its background, structure and aim. Scand J Public Health. 2001;29:300–307.
11. Lauritsen MB, Jørgensen M, Madsen KM, Lemcke S, Toft S, Grove J, et al. Validity of Childhood Autism in the Danish Psychiatric Central Register: Findings from a Cohort Sample Born 1990–1999. J Autism Dev Disord. 2010;40:139–148.
12. Sandin S, Schendel D, Magnusson P, Hultman C, Surén P, Susser E, et al. Autism risk associated with parental age and with increasing difference in age between the parents. Mol Psychiatry. 2016;21:693–700.
13. Hirota T, King BH. Autism Spectrum Disorder: A Review. JAMA. 2023;329:157–168.
14. Chien Y-L, Wu C-S, Chang Y-C, Cheong M-L, Yao T-C, Tsai H-J. Associations between parental psychiatric disorders and autism spectrum disorder in the offspring. Autism Research. 2022;15:2409–2419.
15. Yin W, Pulakka A, Reichenberg A, Kolevzon A, Ludvigsson JF, Risnes K, et al. Association between parental psychiatric disorders and risk of offspring autism spectrum disorder: a Swedish and Finnish population-based cohort study. The Lancet Regional Health - Europe. 2024;40:100902.
16. Crump C, Sundquist J, Sundquist K. PRETERM OR EARLY TERM BIRTH AND RISK OF AUTISM. Pediatrics. 2021;148:e2020032300.
17. Nivins S, Gissler M, Lavebratt C. Neurodevelopmental risks of late-preterm and early-term births: a population-based study from Finland. Bmjph. 2026;4.
18. Carter S, Lin JC, Chow T, Martinez MP, Qiu C, Feldman RK, et al. Preeclampsia Onset, Days to Delivery, and Autism Spectrum Disorders in Offspring: Clinical Birth Cohort Study. JMIR Public Health Surveill. 2024;10:e47396.
19. Mei Z, Yang S, Li L, Zhao K, Zhang H, Liao A, et al. Preeclampsia: The silent catalyst for autism spectrum disorder in offspring. Placenta. 2025. 14 July 2025. https://doi.org/10.1016/j.placenta.2025.07.072.

20. Andersen CH, Thomsen PH, Nohr EA, Lemcke S. Maternal body mass index before pregnancy as a risk factor for ADHD and autism in children. Eur Child Adolesc Psychiatry. 2018;27:139–148.
21. Kim JY, Son MJ, Son CY, Radua J, Eisenhut M, Gressier F, et al. Environmental risk factors and biomarkers for autism spectrum disorder: an umbrella review of the evidence. Lancet Psychiatry. 2019;6:590–600.
22. Hornig M, Bresnahan MA, Che X, Schultz AF, Ukaigwe JE, Eddy ML, et al. Prenatal fever and autism risk. Mol Psychiatry. 2018;23:759–766.
23. Croen LA, Qian Y, Ashwood P, Zerbo O, Schendel D, Pinto-Martin J, et al. Infection and Fever in Pregnancy and Autism Spectrum Disorders: Findings from the Study to Explore Early Development. Autism Res. 2019;12:1551–1561.
24. Jung Y, Lee AM, McKee SA, Picciotto MR. Maternal smoking and autism spectrum disorder: meta-analysis with population smoking metrics as moderators. Sci Rep. 2017;7:4315.
25. Tian LH, Barger B, Pazol K, Schieve LA, Bertrand J, DiGuiseppi C, et al. Prenatal Maternal Alcohol Exposure During the First Trimester of Pregnancy in Relation to Early Learning Ability, Behavioral Problems, and Autistic Traits in Preschool Children With or Without Autism Spectrum Disorder. Autism Research. 2025;18:1087–1096.
26. Isaksson J, Ruchkin V, Ljungström T, Bölte S. Evaluation of Birth Weight and Neurodevelopmental Conditions Among Monozygotic and Dizygotic Twins. JAMA Netw Open. 2023;6:e2321165.
27. Sandin S, Yip BHK, Yin W, Weiss LA, Dougherty JD, Fass S, et al. Examining Sex Differences in Autism Heritability. JAMA Psychiatry. 2024;81:673–680.
28. Bech BH, Nohr EA, Vaeth M, Henriksen TB, Olsen J. Coffee and Fetal Death: A Cohort Study with Prospective Data. Am J Epidemiol. 2005;162:983–990.
29. Dohm-Hansen S, Caruso MG, Nicolas S, Scaife C, O'Leary OF, Nolan YM, et al. Expanding the Proteomics and Metabolomics Toolkit with Methods for Differential Expression Analysis from Transcriptomics. J Proteome Res. 2026;25:1253–1264.
30. Moons KGM, Altman DG, Reitsma JB, Collins GS, Transparent Reporting of a Multivariate Prediction Model for Individual Prognosis or Development Initiative. New Guideline for the Reporting of Studies Developing, Validating, or Updating a Multivariable Clinical Prediction Model: The TRIPOD Statement. Adv Anat Pathol. 2015;22:303–305.
31. Cao X, Tang X, Feng C, Lin J, Zhang H, Liu Q, et al. A Systematic Investigation of Complement and Coagulation-Related Protein in Autism Spectrum Disorder Using Multiple Reaction Monitoring Technology. Neurosci Bull. 2023;39:1623–1637.
32. Noone A, Dowling K, O'Boyle D, Carter M, Golubeva AV, Scaife C, et al. Longitudinal multi-omics analysis of umbilical cord blood and childhood serum in Autism. Mol Psychiatry. 2026;31:701–713.
33. Vacher C-M, Tsompanidis A, Firestein MR, Penn AA. Neuroactive steroid exposure impacts neurodevelopment: Comparison of human and rodent placental contribution. Journal of Neuroendocrinology. 2025;37:e13489.

34. Philosoph LK, Yoffe C, El Biri B, Baryakh P, Saar EG, Mayorov M, et al. Distinct protein profiles in cord blood plasma of children with autism spectrum disorder. Sci Rep. 2025;15:44947.
35. Che X, Hornig M, Bresnahan M, Stoltenberg C, Magnus P, Surén P, et al. Maternal mid-gestational and child cord blood immune signatures are strongly associated with offspring risk of ASD. Mol Psychiatry. 2022;27:1527–1541.
36. Salzbank J, Lacaille H, Gaby J, O'Reilly JJ, Kissner M, Vacher C-M, et al. Microglia alter sex-specific cerebellar myelination following placental hormone loss. Nat Commun. 2025;16:9846.
37. Vacher C-M, Lacaille H, O'Reilly JJ, Salzbank J, Bakalar D, Sebaoui S, et al. Placental endocrine function shapes cerebellar development and social behavior. Nat Neurosci. 2021;24:1392–1401.
38. Baron-Cohen S, Tsompanidis A, Auyeung B, Nørgaard-Pedersen B, Hougaard DM, Abdallah M, et al. Foetal oestrogens and autism. Mol Psychiatry. 2020;25:2970–2978.
39. Baron-Cohen S, Auyeung B, Nørgaard-Pedersen B, Hougaard DM, Abdallah MW, Melgaard L, et al. Elevated fetal steroidogenic activity in autism. Mol Psychiatry. 2015;20:369–376.
40. Kenny L, Hattersley C, Molins B, Buckley C, Povey C, Pellicano E. Which terms should be used to describe autism? Perspectives from the UK autism community. Autism. 2016;20:442–462.
41. Monk R, Whitehouse AJO, Waddington H. The use of language in autism research. Trends in Neurosciences. 2022;45:791–793.
42. Pellicano E, den Houting J. Annual Research Review: Shifting from 'normal science' to neurodiversity in autism science. Journal of Child Psychology and Psychiatry. 2022;63:381–396.
43. AsIAm. As I Am | Ireland's Autism Charity | Resource Library. https://asiam.ie/advice-guidance/resource-library#Autism-and-Language-What-s-the-Right-Word. Accessed 30 May 2026.

Figures

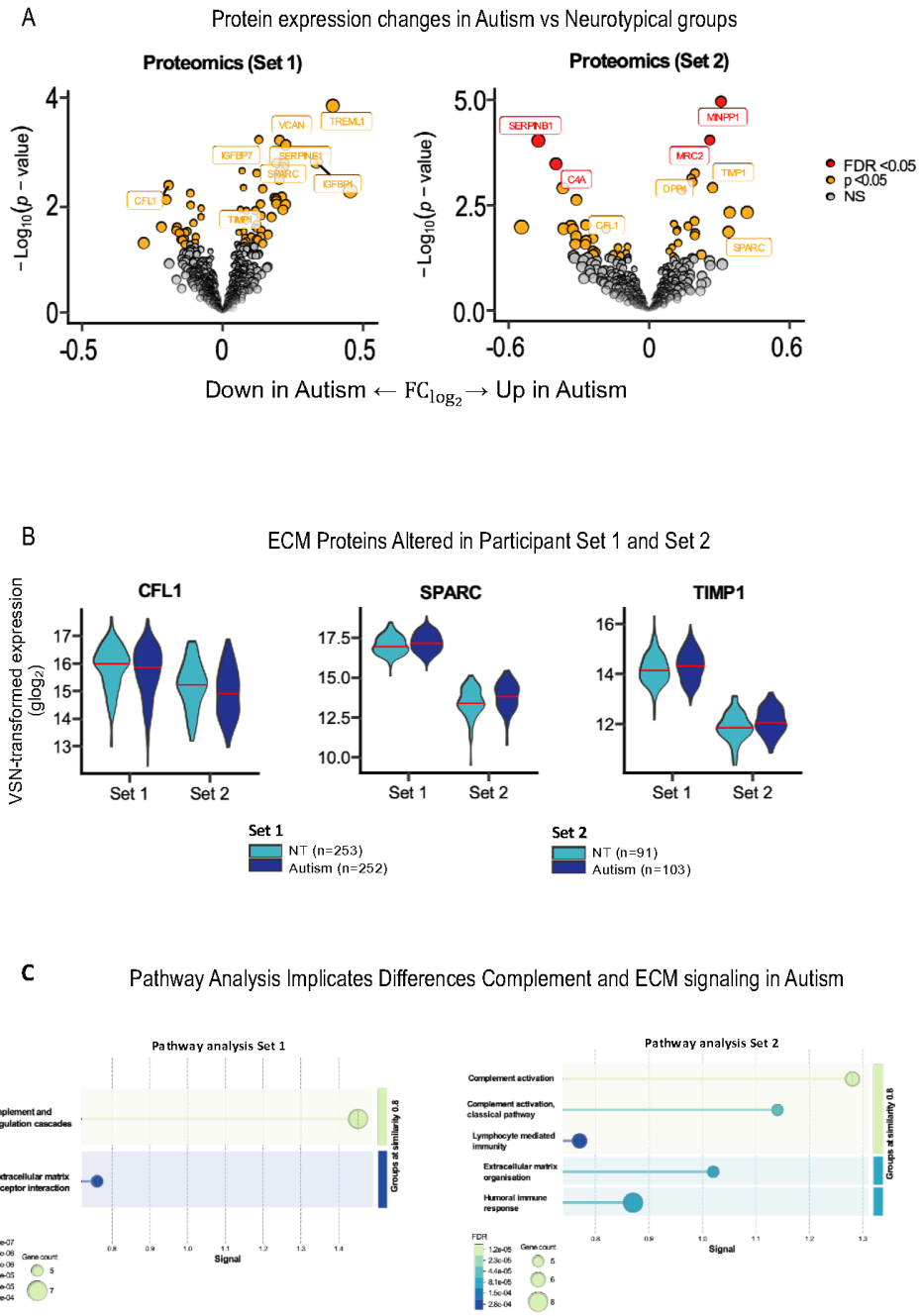

Figure 1

Proteomics Implicates Complement Signalling and Extracellular Matrix (ECM) pathways in Autism

(A) Volcano plots illustrating protein expression fold changes in autism versus neurotypical controls in participant dataset 1 and participant dataset 2. Red dot: FDR<0.05; Yellow: p<0.05; Gray: Not Significant.

(B) Violin plots showing three, nominally significant ($p < 0.05$) proteins (CFL1, SPARC, TIMP1) that were common to both participant dataset 1 and participant dataset 2 and demonstrated similar direction of change. All three proteins are involved in extracellular matrix homeostasis and regulation. These three proteins are depicted in the volcano plots alongside some of the top differentially expressed proteins.

(C) Pathway analysis outputs of the top biological processes implicated in participant dataset 1 and participant dataset 2, showing convergence on complement activation and signalling alongside extracellular matrix (ECM) organisation and remodelling. Participant dataset 2 also illustrates immunity and immune responses as altered in cases versus controls. Signal strength is shown on the horizontal axis, while biological processes are grouped by similarity along the vertical axis.

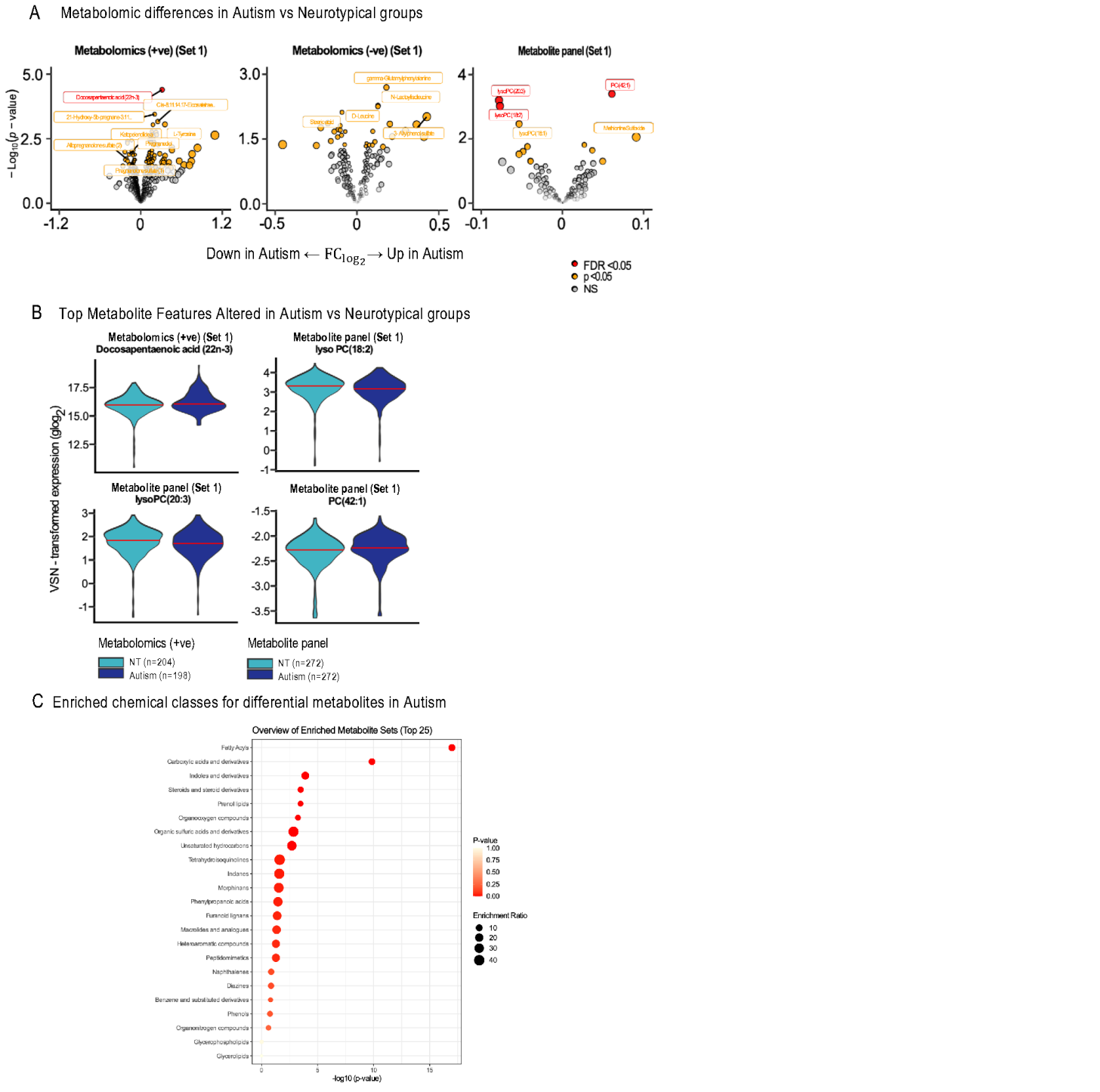

Figure 2

Metabolomic Changes in Unsaturated Fatty Acids and Phospholipids in Autism (Cases) vs Neurotypical (Controls)

(A) Volcano plots from discovery metabolomics analysis in participant dataset 1 illustrate differentially expressed features from LC-MS/MS ve+ and ve- modes (Waters SYNAPT G2-S) and metabolite panel

(Biocrates AbsoluteIDQ® p180). Red dot: FDR<0.05; Yellow: p<0.05; Gray: Not Significant.

(B) Violin plots of differentially expressed metabolites FDR < 0.05.

(C) Enriched chemical classes of nominally differentially expressed (p<0.05) metabolites (n=148) in autism.

Abbreviations: FC (Fold Change); VSN (Variance Stabilizing Normalization); glog (generalized logarithm).

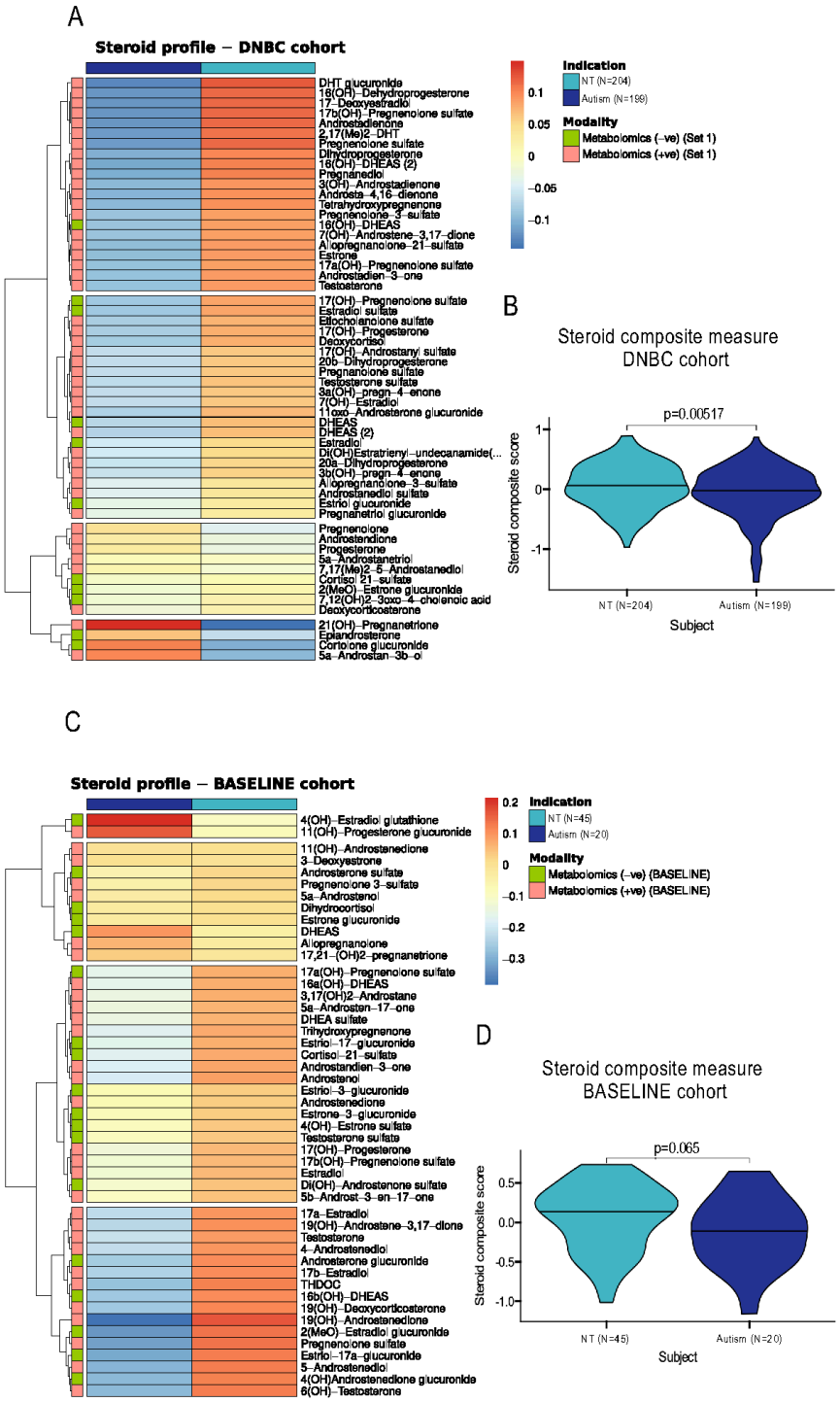

Figure 3

Prenatal Steroid Signature in Autism in DNBC and BASELINE Cohorts

(A) Heatmap showing relative changes in steroids between autism cases and neurotypical controls in the DNBC cohort. Detection modes: positive = green side bar, negative = pink side bar.

(B) Composite steroid score between autism and neurotypical controls in the DNBC cohort (two-tailed, p=0.00517, logistic regression).

(C) Heatmap showing relative changes in steroids between autism cases and neurotypical controls in the BASELINE cohort. Detection modes: positive = green side bar, negative = pink side bar.

(D) Composite steroid score between cases and controls in the BASELINE cohort. The difference of distributions was not significant (one-tailed, p=0.065, Welch's t-test)

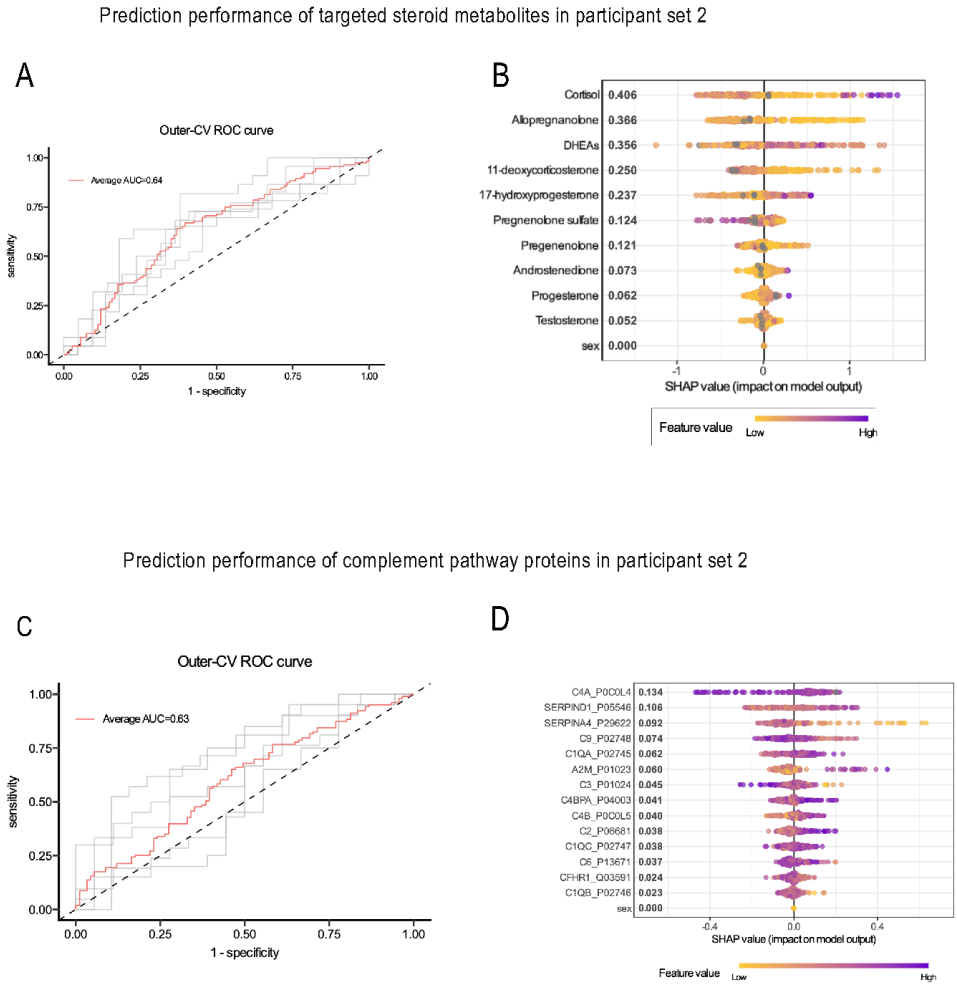

Figure 4

Predictive performance of metabolomic and proteomic features in Autism classification (XGBoost)

(A) ROC curve of outer CV fold performance for targeted steroids model in participant dataset 2. The average area under the ROC across outer folds is indicated.

(B) SHAP plot showing the importance of each steroid feature to the model. The average, unsigned SHAP value across all training examples is indicated on the left (ranked from highest to lowest), while individual, signed average values per training example are plotted.

(C) ROC of outer CV performance for discovery complement and coagulation proteomics model in participant dataset 2 (as in panel A).

(D) SHAP plot of feature importance (as in panel B).

Abbreviations: CV (cross-validation), ROC (Receiver Operating Characteristics), SHAP (SHapley Additive exPlanation)

Supplementary Files

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

- Supplement1.docx
- Supplement2.xlsx
- Supplement3.xlsx