Rule-out test for autism using machine-learning analysis of molecular temporal dynamics in hair - a multicenter study

Vishal Midya[1*], Ghalib A. Bello[2*], Louis A. Gomez[2*], Manuel Ruiz Marin[3], Sujeewa C. Piyankarage[2], Suzy Elhlou[2], Jyoti Chumber[2], Juliet Jaramilo[2], Sophie Dessalle[2], Maayan Yitshak-Sade[1], Alejandra Cantoral[4], Rosalind J. Wright[1,5], Robert Wright[1], Shoji Nakayama[6], Deborah H. Bennett[7#], Rebecca J. Schmidt[7#], Sven Bölte[8-10#], Manish Arora[1,2#]

[1] Dept. of Environmental Medicine, Icahn School of Medicine at Mount Sinai, New York, USA
[2] Linus Biotechnology Inc. North Brunswick, New Jersey, USA
[3]Department of Quantitative Methods, Law and Modern Languages, Universidad Politécnica de Cartagena, Murcia 30201, Spain
[4] Universidad Iberoamericana Ciudad de México, Mexico City, Mexico
[5] Dept. of Public Health, Icahn School of Medicine at Mount Sinai, New York, USA
[6] Exposure Dynamics Research Section, Health and Environmental Risk Division, National Institute for Environmental Studies, Tsukuba, Japan
[7] UC Davis MIND Institute, University of California at Davis, CA, USA
[8] Dept. of Public Health Sciences, School of Medicine, University of California at Davis, CA, USA
[9] Center of Neurodevelopmental Disorders (KIND), Department of Women's and Children's Health, Centre for Psychiatry Research, & Region Stockholm, Stockholm, Sweden
[10]Curtin Autism Research Group, Curtin School of Allied Health, Curtin University, Perth, Australia
*Equal Contribution [#]Equal Contribution

ABSTRACT

BACKGROUND: Early intervention can improve autism-related outcomes. However, no valid biosignature test exists yet for detecting or excluding autism. In addition, most behavior-based assessments of autism are developed for children aged 18 months and older. We developed a hair-strand-based biomarker test (diagnostic aid) to assist clinicians in ruling out autism in children aged 1 month and older.

METHODS: In a multi-national sample of 1697 (from California (two studies, n= 1112), New York City (n= 123), Sweden (n= 306), Japan (n= 110), and Mexico City (n= 46), with 97% below 21 years-of-age), autism was assessed using DSM-5 criteria for autism spectrum disorder or gold standard diagnostic instruments (ADOS-2 and/or ADI-R). A single hair-strand from children collected at 1 month or older was analyzed using laser ablation-inductively coupled plasma-mass spectrometry to sample down the shaft, generating time-series data at a resolution of ~800 timepoints (on average) for 12 elemental intensities. We employed a multi-layered machine learning architecture to leverage the temporality of elemental intensities and optimized the test negative predictive value (NPV) and sensitivity. Models were trained, ensembled, and tuned on participants from California and Sweden, then tested on 580 participants (within-population replication in California and Sweden, and external population testing in New York, Mexico City, and Japan).

RESULTS: The diagnostic aid showed an AUC of 0.75 (95%CI: 0.70-0.79), 97% NPV (95%CI: 0.92-0.99), and 96% sensitivity (95%CI: 0.91-0.98), respectively (estimated prevalence for NPV set at 14%).

Moreover, for those 36 months or younger, the AUC was 0.79 (95%CI: 0.73-0.85), with 97% NPV (95%CI: 0.87-0.99), and 96% sensitivity (95%CI: 0.88-0.99), respectively. Compared to the baseline odds of autism before taking the test (pre-test odds), those who test negative are, on average, ~82% lower in odds of autism diagnosis, whereas those who test positive are ~25% higher.

CONCLUSIONS: By estimating autism likelihood as early as 1 month after birth, early intervention can be delivered with higher precision to young children with developmental support needs. Clinicians may use this diagnostic aid to support early autism diagnosis, improve clinical workflows, and significantly reduce wait times for services.

Keywords: Autism Spectrum disorder; hair-based biomarker; biodynamics; machine learning

INTRODUCTION

Autism is now diagnosed in around 1 to 3% of all children in most mid to high-income countries.[1,2] Due to increasing referral rates and persistent shortages in clinicians specializing in autism diagnosis, wait times for autism diagnostic assessment may be months to years.[3,4] As a result, the median age of diagnosis in the US remains over 4 years, hindering timely intervention. Environmental factors have been shown to moderate or mediate neurodevelopment and autism phenotypes both directly and indirectly through gene-environment interactions.[5-7] Furthermore, diagnosis rates of autism in the US and elsewhere in the world are on the rise (recent data showing 1 in 31 children are diagnosed with autism in the US), and multiple other factors besides genetics, such as environmental exposures, improved screening and diagnostic practices, are likely at play if rates are rising.

Among those environmental factors, exposure to toxic metals and deficiencies in nutritional elements during prenatal and early childhood have been associated with increased likelihood of autism and other developmental conditions.[8,9] Animal studies suggest that elemental exposure can alter brain development by influencing neurotransmission in frontal and subcortical brain regions, some of which are also implicated in autism.[8,10] Elemental biodynamics in humans, i.e., the dynamic processes by which homeostasis of essential and non-essential elements is maintained within the body's systems, are tightly regulated.[11-13] These processes are controlled by intricate regulatory mechanisms that respond to various environmental and internal signals, ensuring that the concentrations of these elements remain within a physiological range.[14,15] Key biological processes (metabolism, acid-base balance, renal function, for example) cyclically regulate the temporal dynamics of elemental stoichiometry to maintain homeostasis.[16-18] Some of these processes are also implicated in changes of amino acid pathways and mitochondrial alterations, and can contribute to autism-related behaviors.[19,20] One hypothesized mechanism underlying divergent neurotypes, therefore, involves elemental homeostasis. The biodynamics of elemental homeostasis guiding early development are tightly controlled by regulatory mechanisms that respond to external environmental perturbations.[21,22] Therefore, biodynamic synchronization in elemental profiles may serve as markers of homeostatic regulation that may be predictive of present and future likelihood of autism-related behaviors.[23]

Fluid-based biomarkers (e.g., blood and urine) provide a "snapshot" of a given time point. Obtaining temporal intensity patterns by collecting repeated blood, urine, or stool samples is generally not feasible due to the labor required. To overcome this challenge, we analyzed elemental intensities in hair growth increments using laser ablation-inductively coupled plasma-mass spectrometry (LA-ICP-MS) to generate time-resolved signals of elemental rhythm at approximately hourly resolution along the shaft, with an average of ~800 sampling points per hair.[24] Since hair grows at ~1 cm a month, these data represent elemental concentration

patterns over time. Using a machine learning (ML) architecture, we leveraged this time-series data to study elemental biodynamics as a biomarker of homeostasis that can distinguish individuals diagnosed with autism from those without. In this study, we develop and validate a novel diagnostic aid that leverages the temporal biodynamics of 12 elements: magnesium (Mg), phosphorus (P), calcium (Ca), copper (Cu), zinc (Zn), strontium (Sr), barium (Ba), lead (Pb), lithium (Li), manganese (Mn), Sulphur (S) and Arsenic (As).

When considering children referred from general pediatricians for autism evaluation by specialists, a significant proportion of children do not ultimately get diagnosed with autism, but all children undergo lengthy behavioral assessments that place demands on clinician time and burden on service users.[25,26] A key missing piece in the autism diagnostic clinical workflow is a non-invasive, objective biochemical marker that can effectively rule out autism and help clinicians prioritize those children at a higher likelihood of autism. Here, we describe a novel machine-learning-driven, hair-strand-based biochemical test that may serve as such an aid for clinicians. This diagnostic aid is designed as a rule-out test for autism in children from as young as 1 month, with high negative predictive values (NPV) and sensitivity. Figure 1 provides an overview of the proposed clinical process and compares it with the current standard of care.

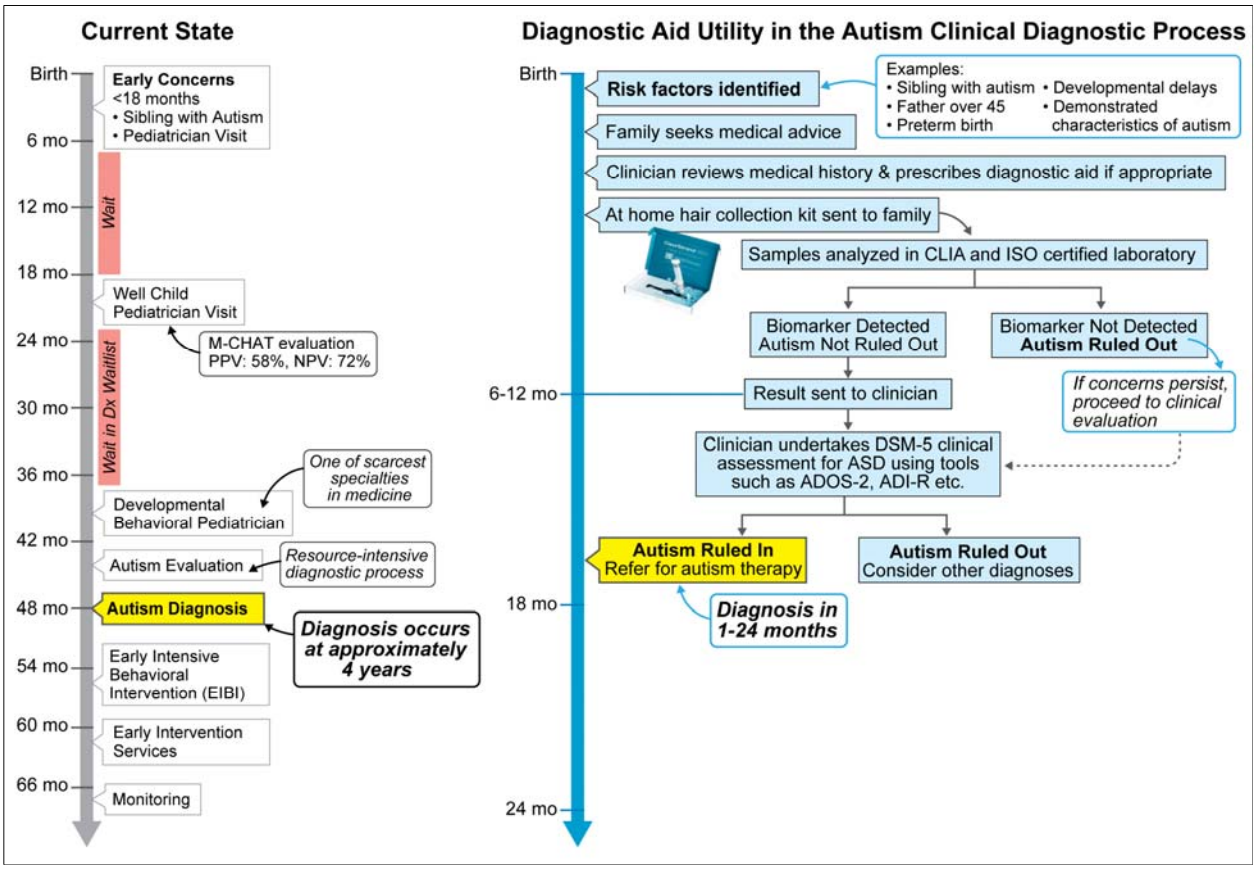

Figure 1. Overview of the current state of guidelines for autism diagnosis and the proposed diagnostic aid-supported workflow. Based on family history, parental concerns, or medical presentation, a clinician can prescribe this diagnostic aid as early as 1 month of age. A negative result indicates that it is unlikely the child has an increased autism likelihood, but parents are nonetheless asked to monitor their child. Children with non-negative results are prioritized for complete clinical assessment. This allows the full diagnostic journey to be completed before the child turns 24 months old. In contrast, under the traditional model, diagnosis occurs at a nationwide median of approximately 4 years. CLIA: Clinical Laboratory

Improvement Amendments; ISO: International Organization for Standardization; DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; M-CHAT: Modified Checklist for Autism in Toddlers.

MATERIALS AND METHODS

STUDY POPULATIONS

The combined sample (N = 1697) used in this analysis was collected from diverse study populations across multiple countries, including the United States, Sweden, Mexico, and Japan. Details on the IRB can be found in the appendix, p2.

CHARGE and MARBLES (California, USA)

Data were collected from the Childhood Autism Risks from Genetics and Environment (CHARGE) and the Markers of Autism Risk in Babies—Learning Early Signs (MARBLES) studies, recruited in California by the MIND Institute at the University of California, Davis. [27,28] Informed consent was obtained from each study participant before participation and data collection in both studies. The protocols for the CHARGE and MARBLES cohorts have been previously described.[27,28] See appendix p2 for additional details. CHARGE began recruitment in 2003. It is a case-control study that includes children receiving services for autism or developmental delay from Regional Centers that contract with the California Department of Developmental Services and children recruited from the general population from birth records, frequency-matched to the autism cases on sex, geographic region, and age. Those entering with autism had the diagnosis confirmed by both the Autism Diagnostic Observation Schedule Second Edition (ADOS-2) and Autism Diagnostic Interview—Revised (ADI-R). All other participants were screened with the Social Communication Questionnaire (SCQ).[29] Those children whose SCQ score was above 14 were further administered the ADI-R and ADOS-2.[27] CHARGE samples used in this study include either hair saved by parents from the child's first haircut or hair collected at the child's clinic visit; thus, some children were diagnosed after, and some before, the hair sample collection. For CHARGE, the non-autism group included all children who did not meet the diagnostic criteria for autism, encompassing those classified as typically developing, developmentally delayed, with autism symptoms but not meeting the criteria for autism diagnosis, or with early concerns (ADOS-2 total score ≥ 7). MARBLES is an elevated-likelihood birth cohort recruiting pregnant women who either have a child or a first-degree relative with autism. Recruitment began in 2006. Children were evaluated at 3 years of age with ADOS-2. For MARBLES samples, hair was typically collected between 3 and 24 months during a clinic visit, and an autism diagnosis was made around age 36 months. For MARBLES, the non-autism group included all children who did not meet the diagnostic criteria for autism, including those classified as typically developing and those classified as not typically developing (e.g., developmentally delayed). Clinicians administering the CHARGE and MARBLES instruments are trained and have formally demonstrated research reliability on the ADOS. Both CHARGE and MARBLES also assessed children's cognitive development using the Mullen Scales of Early Learning (MSEL).[30]

Seaver and PRISM (New York, USA)

Data was collected from the Seaver Autism Center at the Icahn School of Medicine at Mount Sinai, New York City.[31] This center receives referrals related to autism from the surrounding area. Beginning in 2016, families already engaged with Seaver Autism Center were approached if they had a child with a confirmed autism diagnosis and at least one sibling without autism, with hair samples typically collected from both siblings. Diagnoses were supported by the

ADOS-2 and ADI-R, along with a clinical evaluation conducted by a board-certified child and adolescent psychiatrist or a licensed clinical psychologist, in accordance with the DSM-5 diagnostic criteria. We further included general population-based participants drawn from the PRogramming of Intergenerational Stress Mechanisms (PRISM) study, which recruited children whose mothers received prenatal care at Mount Sinai, into a longitudinal study designed to investigate the influence of psychological stress and other environmental influences on child development.[32] The child version of the Autism Spectrum Quotient (AQ-10) was completed for all children, and no children were referred for a specialist diagnostic assessment. In this paper, these children were treated as non-autism participants.[33]

RATSS (Sweden)

Data was also collected from the Roots of Autism and ADHD Twin Study Sweden (RATSS).[34] Participants (twins) were recruited for this study between 2011 and 2016. Participants included both concordant and discordant twin pairs: those in which one or both twins met the criteria for autism, as well as pairs in which neither twin did. Details on participant recruitment have been previously published.[35] Autism screening was conducted at age nine. DSM-5 criteria were used for autism diagnosis, and results were corroborated using the ADI-R and ADOS-2. RATSS included detailed data on ADHD and intellectual disability diagnoses, as well as psychiatric comorbidities. Given the availability of this data, participants with any intellectual disability or other psychiatric comorbidities were excluded, since children with autism have a higher likelihood of being misdiagnosed with intellectual disability.[36,37] For RATSS, the non-autism group comprised participants who did not meet the diagnostic criteria for autism, whether typically developing or with other neurodevelopmental conditions.

JECS (Japan)

Data were additionally obtained from the Japan Environment and Children's Study (JECS), a large-scale, nationwide birth cohort designed to investigate environmental influences on children's health.[38] Recruitment was conducted across 15 regional centers throughout Japan. Hair samples were collected from infants at approximately one month of age and archived at the JECS biorepository under standardized conditions. Autism diagnosis was carried out around age four. For the present study, 220 children were randomly selected from among participants who had completed the three-year follow-up questionnaire. Of these, only 110 participants were treated as non-autism samples (using AQ-10) and included in the validation set (to determine whether including a geographically diverse set of non-autism samples affected the accuracy of our models). Autistic individuals identified within JECS were excluded due to concerns about diagnostic reliability, particularly the lack of DSM-5 consensus with ADI-R and ADOS-2. While this precluded balanced autism/non-autism analyses within JECS, inclusion of non-autism samples broadened the demographic and geographic scope of the validation dataset.

Mexico City (Mexico)

This data comes from Mexico City. Recruitment was conducted through a campaign by Domus Institute of Autism and Iberoamericana University, focusing on families with undiagnosed children younger than 7 years, and with the following inclusion criteria: children with developmental delays or signs of autism, or children with a sibling or immediate family member diagnosed with autism. Children accompanied by their parents provided signed informed consent and hair specimens at the study visit (Domus Institute of Autism). Autism diagnostic status was determined based on clinical evaluations conducted by trained and standardized psychologists within the same week as hair collection from August 2024 to May 2025. During the evaluation, clinicians systematically coded and scored observed behaviors according to standardized ADOS-2 protocols. Importantly, the ADOS-2 was not used as a stand-alone diagnostic tool; its results were integrated with developmental history, family interviews, clinical

observation, and complementary assessments (including ADI-R and cognitive and adaptive evaluations) to reach a reliable and comprehensive clinical diagnosis of autism. For this cohort, the non-autism group consisted of participants who did not meet the diagnostic criteria for autism.

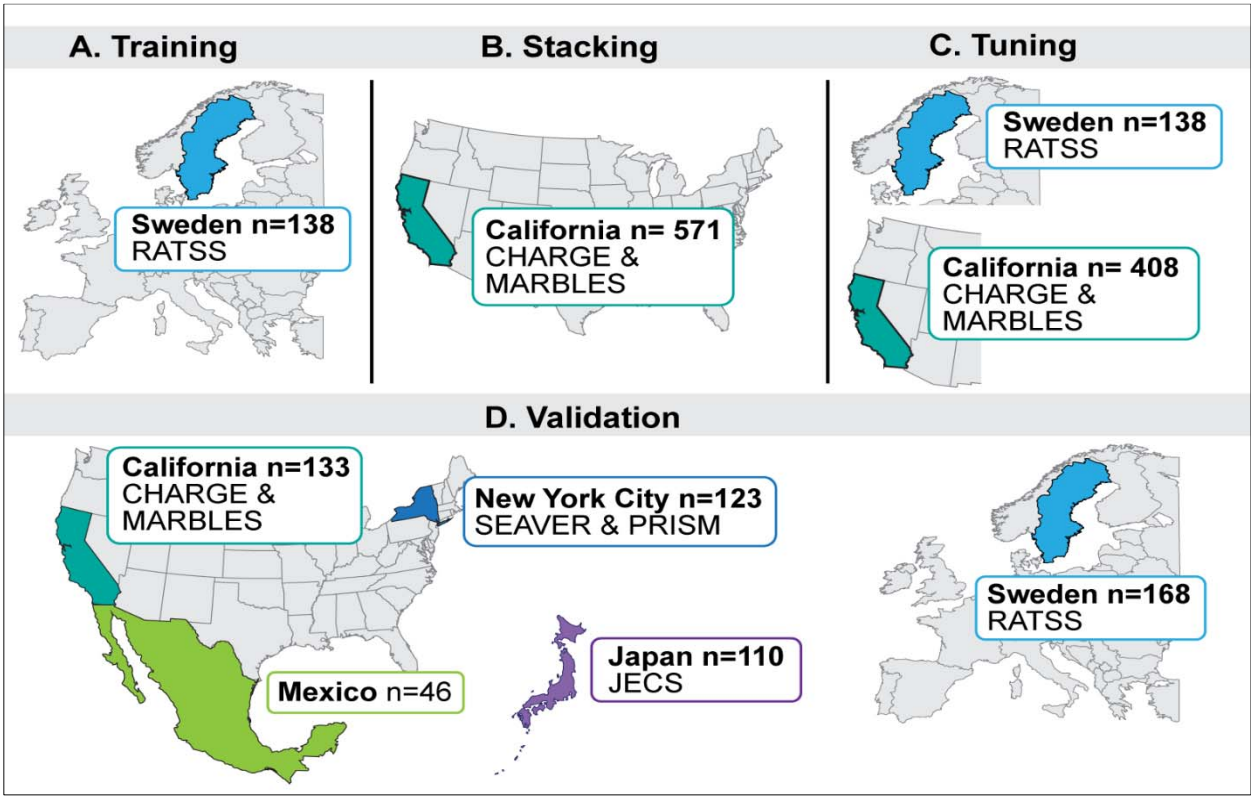

Figure 2: Overall schematic of cohorts used in training, stacking, tuning, and validation tasks. In training, data from the RATSS cohort (n = 138, Sweden) were used. During the stacking phase, data from the CHARGE study and MARBLES cohort (n = 571, California, USA) were used. The threshold for the final model was tuned using the same RATSS cohort (n = 138) used for training, as well as additional CHARGE and MARBLES cohorts (n = 408). Note that the tuning data from the CHARGE and MARBLES cohorts included only non-autism samples, ensuring the model observes variations in non-autism samples and is therefore tuned as a rule-out test. Lastly, the final model was tested on additional CHARGE, MARBLES, and RATSS data, as well as external validation data from Mexico City (n = 46), New York City (n = 123), and Japan (n = 110).

LABORATORY PROTOCOLS AND SIGNAL PROCESSING

Elemental measurements were acquired by LA-ICP-MS (laser ablation-inductively coupled plasma-mass spectrometry) under a standardized daily tuning and ongoing QC regime, using matrix-matched standards and certified reference materials. Runs were released only after meeting predefined QA/QC criteria. Full details on calibration, tuning targets, QC schedules, control charting, and other QA/QC procedures are provided in the supplementary section (appendix p 1-2). Raw ion counts for each element per ablation time point are used to construct the elemental time series. To correct for hair-density/ablation variability and instrumental drift, we applied pointwise sulfur (S) normalization: at each time point, we computed the ratio of the element count to the Sulfur count (e.g., Cu/S). Sulfur itself was not normalized. These ratios were used for all downstream feature extraction and modelling across 12 elements: magnesium (Mg), phosphorus (P), calcium (Ca), copper (Cu), zinc (Zn), strontium (Sr), barium (Ba), lead (Pb), lithium (Li), manganese (Mn), Sulphur (S) and Arsenic (As).

STATISTICAL METHODS

On a sample of individuals drawn from the RATSS, CHARGE, and MARBLES study cohorts, we trained (and stacked) ML models using temporal biodynamic features computed from multivariate elemental time series. Modeling in this context was challenging due to potential heterogeneity in autism etiologies, which may manifest as differences in elemental biodynamics.[39,40] In order to capture diverse aspects of elemental biodynamics, we used 8 different feature extraction techniques, encompassing both univariate measures (e.g., entropy-complexity) computed for individual elements and bivariate measures (e.g., recurrence quantification analysis and causal network centrality), computed across elements pairs.[41-43] To account for potential heterogeneity in autism, we stratified the models by sex. Although the biological mechanisms underlying sex differences in the etiology of autism are not fully understood, prior research suggests different phenotypic signatures.[44-47] This stratified approach enabled modeling within more clinically homogeneous subgroups.

Modelling Approach

We developed a framework that allowed us to consider multiple model configurations while ensuring that the selected models for validation met our *a priori* selection criteria. Our framework, termed "StrandML," is run separately for each of the demographic subgroups mentioned above. The StrandML framework (see Figure 3) is structured into two primary phases: a Training Phase, which involves model development and statistical testing to filter promising configurations, and a Stacking Phase, where these filtered configurations are combined into a model ensemble using stacked generalization.[48] A final, formal validation was conducted on hold-out data exclusively for performance reporting of the selected models. Here we provide a brief overview. See supplementary method for more description (appendix p3-5).

Training Phase: We constructed a candidate model space by exhaustively combining two components: (1) predefined elemental subsets, including all 55 unique pairs among 11 elements and a full model using all 11 elements; (2) 8 time series feature extraction techniques. This resulted in 448 candidate models per demographic subgroup. For each element-feature combination, we fit an elastic-net-penalized logistic regression to the derived features. Performance was evaluated using repeated stratified cross-validation. The mean cross-validated AUC was computed for each model (AUC^{tr}), and only those with $AUC^{tr}>0.5$ were advanced to the stacking phase.

Stacking Phase: Candidate models that passed training-phase criteria were combined via stacked generalization. An independent stacking set was used to fit a meta-learner that ingests the base learners' predicted probabilities as features. For the meta-learner, we used an *ExtraTrees* classifier (also called Extremely Randomized Trees).[49] To assess the relative contribution of the individual base models within the stacked ensemble, we applied SHAP analyses (tree-SHAP) to the stacking dataset, attributing the meta-learner's predictions to each base-model feature.[50] The stacking process was carried out within each subgroup/sex.

Threshold Tuning: StrandML is designed as a rule-out test; therefore, our goal was to identify probability thresholds that ensure high sensitivity and NPV (both > 90%) in the final testing stage. Once the fitted meta-learner (from the stacking phase) was applied to the tuning set and predicted probabilities were generated, we searched for a probability threshold that yielded sensitivity and specificity matching our target performance. Typically, a range of probability values meets the target criteria. In this work, we use the 10th percentile probability threshold and conduct a sensitivity analysis to assess the robustness of other values within the range.

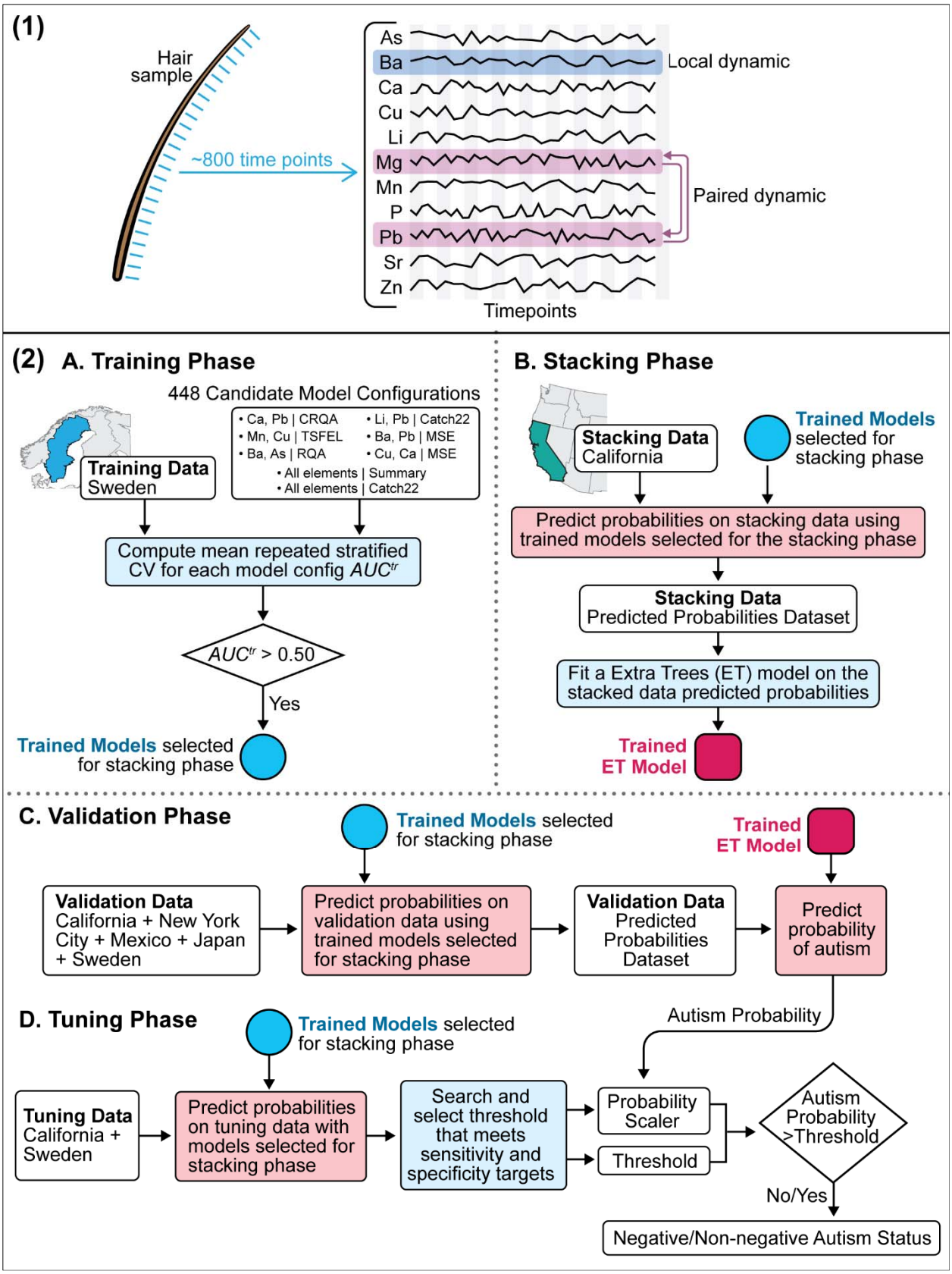
(1)
Hair sample
~800 time points
As
Ba
Ca
Cu
Li
Mg
Mn
P
Pb
Sr
Zn
Local dynamic
Paired dynamic
Timepoints
(2) A. Training Phase
448 Candidate Model Configurations
• Ca, Pb | CRQA
• Mn, Cu | TSFEL
• Ba, As | RQA
• Li, Pb | Catch22
• Ba, Pb | MSE
• Cu, Ca | MSE
• All elements | Summary
• All elements | Catch22
Training Data Sweden
Compute mean repeated stratified CV for each model config AUCtr
AUCtr > 0.50
Yes
Trained Models selected for stacking phase
B. Stacking Phase
Stacking Data California
Trained Models selected for stacking phase
Predict probabilities on stacking data using trained models selected for the stacking phase
Stacking Data Predicted Probabilities Dataset
Fit a Extra Trees (ET) model on the stacked data predicted probabilities
Trained ET Model
C. Validation Phase
Trained Models selected for stacking phase
Trained ET Model
Validation Data California + New York City + Mexico + Japan + Sweden
Predict probabilities on validation data using trained models selected for stacking phase
Validation Data Predicted Probabilities Dataset
Predict probability of autism
Autism Probability
D. Tuning Phase
Trained Models selected for stacking phase
Tuning Data California + Sweden
Predict probabilities on tuning data with models selected for stacking phase
Search and select threshold that meets sensitivity and specificity targets
Probability Scaler
Threshold
Autism Probability >Threshold
No/Yes
Negative/Non-negative Autism Status

Figure 3: Figure presents an overview of (1) the temporality of elemental intensities from a hair strand, and (2) our modeling approach, StrandML. Our machine-learning models explored both local and pair-wise dynamics in elemental intensities. (2A) denotes the training phase where models were trained using data from the RATSS cohort (Sweden), (2B) denotes the stacking phases where chosen models were stacked based on the CHARGE & MARBLES cohorts (California, USA), (2C-2D) denotes the validation phase where the final model is tested on multiple different cohorts, and once the model is finalized and tested on the multiple cohorts, threshold tuning is performed by optimizing sensitivity and NPV to ensure the final model can be utilized as a rule-out test. ET: Extra Trees Classifier. AUC^{tr}: training AUC

Final Testing and Evaluation Metrics: The final models selected for each demographic subgroup underwent a formal validation process, using an aggregate of four independent datasets that had not been used for model training or stacking. These datasets included additional samples from the CHARGE and MARBLES cohorts, as well as four external cohorts: SEAVER and PRISM (New York, USA), samples from Mexico City, the JECS cohort from Japan, and the RATSS cohort in Sweden. The inclusion of these geographically and demographically distinct populations enabled a robust assessment of the models' generalizability to new, independent samples. Model performance was primarily assessed using AUC. We obtained permutation-based p-values for the AUC using 5000 random-label permutations, and 95% confidence intervals were estimated using bias-corrected accelerated bootstrap resampling with 1000 iterations. In addition to AUC, we evaluated a set of threshold-dependent performance metrics: sensitivity, specificity, and balanced accuracy. Additionally, the positive and negative predictive values (PPV and NPV) were estimated under a prevalence of 14% for elevated risk.[51,52] To enhance clinical interpretability, we also computed positive and negative diagnostic likelihood ratios (DLRs).

Training, Stacking, Tuning, and Validation Subsets: For model training, 138 RATSS samples were used. A further 571 samples (339 from CHARGE and 232 from MARBLES) were allocated for stacking. An additional 546 samples were used as a tuning set, comprising 408 non-autism samples from CHARGE and MARBLES, and the 138 RATSS samples used in the model training phase. Finally, 580 samples were reserved for independent validation. These samples were collected from multiple cohorts to provide geographic and demographic diversity, including CHARGE/MARBLES (California), RATSS (Sweden), Mexico, JECS (Japan), and New York (Seaver and PRISM). Figure 2 shows the schematic of cohorts at various modeling stages and validation. Subjects were stratified by sex, as discussed in the Methods section. These subgroups varied in size and prevalence of autism (Table S1, appendix p6). To evaluate the robustness of our modelling framework, we performed multiple sensitivity analyses (see appendix p5 for full details).

RESULTS

Sample Characteristics

A total of 1697 samples were included across six cohorts: CHARGE (n = 825) and MARBLES (n = 287) from California, RATSS from Sweden (n = 306), Seaver (n = 39) and PRISM (n = 84) cohorts from New York, JECS from Japan (n = 110), and a cohort from Mexico (n = 46). Only one hair strand was used for each sample. Table 1 summarizes the demographic and elemental characteristics within each cohort. In CHARGE, males comprised 77.1% of all samples, with a median age at the time of hair collection of 36 months. In MARBLES, participants were younger (median 12.5 months) at the time of hair collection, with a more balanced sex distribution overall (55.8% males), although cases were disproportionately male (74.2%). Among the non-autism group, 51% and 20% had developmental delays, in CHARGE and MARBLES, respectively. RATSS participants were older at the time of hair collection (median 168 months), with 55.6%

males. The Seaver cohort had a median age of 72 months and also showed a predominance of males among cases. The Mexico cohort had a median age of 35.5 months at the time of hair collection and was predominantly male (73.9%). The PRISM cohort consisted of younger participants (compared to the other NYC cohort), with a median age of 48 months at the time of hair collection, with 61% males. Overall, the youngest participants were from JECS (all non-autism), with a median age of hair collection of 1 month and a balanced sex distribution (50% males). Average ablated hair lengths were highly consistent across CHARGE, MARBLES, and Seaver (~1.19 cm), somewhat lower in PRISM and Mexico (~1.16 cm) and JECS (~1.14 cm), and substantially shorter in RATSS (~0.91 cm). Full graphical comparisons of measured elements across subgroups are provided in Figure S2-S13 (appendix p10-21).

Table 1: Demographic and elemental statistics for all samples included in the study

	All (n=1697)	**Autism (n=354)**	**Non-autism (n=1343)**	**p-value***
CHARGE (California, US, n=825, Autism=164, Non-Autism=661)				
Age (in months) at hair collection, *median (IQR)*	36 (20)	27 (20)	41 (19)	<0.01
Sex, *n (%)*				<0.01
Female	189 (22.91%)	17 (10.37%)	172 (26.02%)	
Male	636 (77.09%)	147 (89.63%)	489 (73.98%)	
Hair length (in centimeters), *Mean (SD)*	1.19 (0.01)	1.19 (0.01)	1.19 (0.01)	0.12
Hair length (in days of growth), *Mean (SD)*	35.72 (0.43)	35.76 (0.39)	35.71 (0.44)	0.12
MARBLES (California, US, n=287, Autism=66, Non-autism=221)				
Age (in months) at hair collection, *median (IQR)*	12.5 (11.7)	12.7 (12.0)	12.5 (11.1)	0.23
Sex, *n (%)*				< 0.01
Female	127 (44.3%)	17 (25.7%)	110 (49.8%)	
Male	160 (55.8%)	49 (74.2%)	111 (50.2%)	
Hair length (in centimeters), *Mean (SD)*	1.19 (0.02)	1.19 (0.02)	1.19 (0.02)	0.36
Hair length (in days of growth), *Mean (SD)*	35.64 (0.47)	35.65 (0.49)	35.64 (0.47)	0.36
RATSS (Sweden, n=306, Autism=70, Non-autism=236)				
Age (in months) at hair collection in months, *median (IQR)*	168 (84)	156.0 (48)	186.0 (84)	<0.01
Sex, *n (%)*				<0.01
Female	136 (44.44%)	21 (30.00%)	115 (48.73%)	

Male	170 (55.56%)	49 (70.00%)	121 (51.27%)	
Hair length (in centimeters), *Mean (SD)*	0.91 (0.29)	0.80 (0.31)	0.94 (0.28)	<0.01
Hair length (in days of growth), *Mean (SD)*	27.16 (8.84)	24.13 (9.22)	28.07 (8.54)	<0.01
SEAVER (New York City, US, n=39, Autism=21, Non-autism=18)				
Age (in months) at hair collection, *median (IQR)*	72 (54)	72 (72)	90 (48)	0.81
Sex, *n (%)*				<0.01
Female	14 (35.90%)	3 (14.29%)	11 (61.11%)	
Male	25 (64.10%)	18 (85.71%)	7 (38.89%)	
Hair length (in centimeters), *Mean (SD)*	1.19 (0.01)	1.19 (0.00)	1.19 (0.01)	0.38
Hair length (in days of growth), *Mean (SD)*	35.70 (0.17)	35.70 (0.07)	35.71 (0.24)	0.38
Mexico (Mexico City, n=46, Autism=33, Non-autism=13)				
Age (in months) at hair collection, *median (IQR)*	35.5 (22.75)	38.0 (25.0)	30.0 (19.0)	0.03
Sex, *n (%)*				<0.01
Female	12 (26.09%)	4 (12.12%)	8 (61.54%)	
Male	34 (73.91%)	29 (87.88%)	5 (38.46%)	
Hair length (in centimeters), *Mean (SD)*	1.16 (0.02)	1.17 (0.01)	1.15 (0.02)	0.01
Hair length (in days of growth), *Mean (SD)*	34.82 (0.58)	34.98 (0.45)	34.40 (0.66)	0.01
PRISM (New York City, US, n=84, Autism=0, Non-autism=84)				
Age (in months) at hair collection, *median (IQR)*	48.0 (7.25)	-	48.0 (7.25)	-
Sex, *n (%)*				
Female	33 (39.29%)	-	33 (39.29%)	-
Male	51 (60.71%)	-	51 (60.71%)	-
Hair length (in centimeters), *Mean (SD)*	1.16 (0.01)	-	1.16 (0.01)	-
Hair length (in days of growth), *Mean (SD)*	34.75 (2.87)	-	34.75 (2.87)	-

JECS (Japan, n=110, Autism=0, Non-autism=110)				
Age (in months) at hair collection, *median (IQR)*	1.0 (0.0)		1.0 (0.0)	-
Sex, *n (%)*				
Female	55 (50%)		55 (50%)	
Male	55 (50%)		55 (50%)	
Hair length (in centimeters), *Mean (SD)*	1.14 (0.11)	-	1.14 (0.11)	-
Hair length (in days of growth), *Mean (SD)*	34.16 (3.20)	-	34.16 (3.20)	-

*The p-value was calculated using the chi-squared test for categorical variables (except in cases of low cell counts, in which case Fisher's Exact test was used) and the Mann-Whitney U test for continuous variables.

Stacking Stage

Base learners that met the training-phase criterion ($AUC^{tr} > 0.5$) were advanced to the stacking stage, yielding 355 models for males and 305 for females. SHAP analyses of the meta-learner on the stacking set indicated distributed contributions rather than reliance on any single base model. In females, the highest-weighted contributors involved Li-Zn and Li-Ba dynamics (using causal centrality features). In males, contributions were particularly diffuse, with prominent pairs such as P-Ba (using multiscale cross-entropy features) and Mg-Cu and Mg-Zn (using Catch22 features). As a robustness check, we replaced the Extra Trees stacking algorithm with a support vector machine (SVM); the female importance profile was largely concordant (especially among top contributors), but concordance was weaker in males, consistent with the broader dispersion of signals in that subgroup.

Validation Performance

The sex-specific ensembles were then evaluated on independent validation cohorts comprising 580 samples from CHARGE/MARBLES (California), RATSS (Sweden), Mexico, JECS (Japan), and New York (Seaver and PRISM). The age range across these samples was approximately 1 month to 444 months (~37 years). Both PRISM and JECS contributed only non-autism samples, as autism cases were either unavailable or excluded. This evaluation provided an unbiased assessment of generalizability across geographically and demographically diverse populations. The results are summarized in Table 2. Overall, the model achieved an AUC of 0.75 (95% CI: 0.70-0.79, permutation-based p-value < 0.001), with a sensitivity of 96% (95% CI: 0.91-0.98), an NPV (at 14% prevalence) of 97% (95% CI: 0.92-0.99), and a DLR- of 0.18. Note that the overall AUC is influenced by the high predicted probabilities of male cases and the lower predicted probabilities of the female non-autism samples. We estimated the overall AUC directly from predicted probabilities pooled across sexes (as opposed to weighted averages of sex-specific AUCs) for the following reasons: (1) the ease of thresholding for the rule-out test, and (2) most

Group	Subset	n	Autis	Non-	AUC	NPV†	PPV†	Sensitivit	Specificit	DLR	DL

cohorts are not sufficiently balanced by sex to consider sex-specific weighting. The overall performance indicates that the model performed well as a rule-out test, minimizing false negatives while maintaining a high probability that those classified as ‘non-negative’ are truly unaffected.

Performance was broadly consistent across geographic cohorts with no evidence for a cohort effect. Within the US, AUCs ranged from 0.74 for New York to 0.77 for California. Non-US cohorts showed similar stability, with Mexico (AUC = 0.78) and Sweden (AUC = 0.71) performing in the same range. Age-stratified analyses showed that performance was strongest in younger children, though class imbalance was more pronounced. Among children aged 36 months or less, the model achieved an AUC of 0.79, a sensitivity of 96%, and an NPV of 97%. In the younger subgroups (children aged 18 months or less, 12 months or less, and 6 months or less), AUC was 0.79, 0.79, and 0.78, respectively, with sensitivity and NPV estimated at 100% in the two youngest subgroups. Note that these 100% estimates reflect the absence of false negatives in this sample and should be interpreted with caution, given the modest sample size and number of cases. In addition, our sensitivity analyses confirmed that our modelling framework maintains consistent performance across multiple robustness evaluations. See appendix p5 for full details.

Table 2: Diagnostic aid rule-out performance metrics

			m	autism				y	y	+	R-
Age* ≤ 6 months		126	5	121	0.78	1.00	0.16	1.00	0.17	1.20	0.00
Age* ≤ 12 months		145	10	135	0.79	1.00	0.16	1.00	0.17	1.21	0.00
Age* ≤ 18 months		179	16	163	0.79	0.95	0.16	0.94	0.18	1.15	0.34
Age* ≤ 36 months		279	56	223	0.79	0.97	0.16	0.96	0.18	1.18	0.19
Age* ≤ 60 months		384	74	310	0.78	0.98	0.17	0.97	0.21	1.23	0.13
Age* ≤ 120 months		426	93	333	0.73	0.96	0.16	0.95	0.21	1.20	0.26
Sex	Female	220	12	208	0.69	0.96	0.18	0.92	0.34	1.39	0.24
	Male	360	107	253	0.64	0.96	0.16	0.96	0.15	1.13	0.26
All US	Overall	256	58	198	0.76	0.98	0.17	0.97	0.24	1.27	0.15
	California (CHARGE)	80	31	49	0.76	0.95	0.16	0.94	0.18	1.15	0.35
	California (MARBLES)	53	6	47	0.79	1.00	0.18	1.00	0.26	1.34	0.00
	New York (SEAVER & PRISM)[#]	123	21	102	0.74	1.00	0.18	1.00	0.25	1.34	0.00
All US & Japan (JECS)[††]		366	58	308	0.77	0.97	0.17	0.97	0.21	1.23	0.16
Mexico		46	33	13	0.78	0.94	0.15	0.97	0.08	1.05	0.39
Sweden (RATSS)***		168	28	140	0.71	0.96	0.18	0.93	0.29	1.31	0.24
All combined[†††]		580	119	461	0.75**	0.97	0.17	0.96	0.23	1.25	0.18

*** excluding intellectual disability and other psychiatric diagnoses (clinical diagnosis of intellectual disability was only available in this cohort[34]; however, the Mullen early learning composite score was available for both CHARGE and MARBLES, see sensitivity analysis).

*Age refers to age at hair collection

**Overall AUC is influenced by the high predicted probability of male autism cases and the low predicted probability of female non-autism samples. We estimated the overall AUC directly from predicted probabilities (rather than weighted averages) to facilitate thresholding for the rule-out test and because most cohorts are not sufficiently well-balanced between males and females to justify sex-specific weighting.

[#]PRISM is all non-autism samples, so AUC could not be calculated just for PRISM, therefore we combined PRISM with the other NYC Cohort

[†]NPV and PPV are adjusted for an assumed 14% autism prevalence (elevated-risk population)

[††]Only non-autistic participants were considered for JECS to illustrate the model's capability to accommodate varied geographical samples

[†††]All US & Japan (JECS), Mexico, and Sweden (RATSS) were considered in the combined sample of 580

Discussion

Here, we introduced a novel hair-based biochemical assay designed to rule-out autism from as early as age 1 month. The diagnostic aid is intended to be used alongside other clinical evaluations, particularly in children between 1 and 36 months of age with increased likelihood or suspicion of autism (e.g., siblings of individuals with autism, children with developmental delays, or children presenting early signs of autism). The purpose of the diagnostic aid is to assist clinicians in efficiently excluding autism in the differential diagnostic process, reducing wait times for referrals while maintaining high safety against false negatives. For a rule-out test, the key performance criteria are high sensitivity and a high NPV. Across multiple cohorts, the diagnostic aid achieved a high sensitivity at 96% overall and an NPV of approximately 97% in an elevated-risk population (with an assumed 14% rate of autism) and over 99% in the general population (with a 3.2% rate of autism).[2,51] This means that a negative test result provides strong reassurance that a child does not fulfill the diagnostic criteria for autism. Diagnostic likelihood ratios reinforce this interpretation, with a DLR- of 0.18 indicating an 82% reduction in autism likelihood relative to the background prevalence. These metrics suggest that the assay could enable healthcare providers to efficiently rule out autism-negative individuals on long waiting lists, thereby preserving clinician time for those more likely to be on the autism spectrum and improving timely access to clinical services and reducing the burden for service users. Moreover, in mid to low-income countries, this test could help to shorten the time for diagnoses, since there is a shortage of economic resources and a lack of knowledge about autism that can delay early detection. To the best of our knowledge, this diagnostic aid is currently the only biochemical rule-out test for autism that can be implemented in the first year of life. Furthermore, the non-invasive nature of hair sampling enhances the assay's clinical utility, facilitating broader accessibility and compliance.

Multiple potential biomarkers of autism are currently being extensively studied, ranging from genetic (autism polygenic score), epigenetic (histone modification changes), immunologic (brain autoantibodies, folate receptor alpha autoantibody, and cytokines), metabolic (altered purine metabolism, ATP-related purinergic signaling, fatty acid oxidation, as well as change in metabolic pathways related to energy, neurotransmitters, and amino acids, Mitochondrial dysfunction), neurological (Morphology and diffusion tensor imaging, MRI, visual attention), and autonomic (Heart rate variability, dentine-based marker).[53-66] Although biomarker research on autism is quite active, most promising biomarkers are still in their early developmental stages. Furthermore, most are not yet suitable for large-scale clinical implementation due to either their potential invasive nature or the difficulty of implementation. Almost all studies on the effectiveness of biomarkers are based on either correlational or association-type studies; therefore, although such results can yield robust statistical inferences, generalizability and the

identification of predictive patterns remain challenging.[67] Finally, the plausibility of varied autism subtypes renders such association-based results difficult to generalize to external, unseen samples.

This work builds on our previous proof-of-concept experiment, which identified a few features with a high degree of association using a much smaller sample.[68] However, a strong association does not necessarily imply a firm prediction.[69] In this paper, we implemented a rigorous ML architecture using a much larger dataset to ensure that the results obtained can generalize to unseen validation data. We leveraged state-of-the-art techniques to model elemental biodynamic features, potentially capturing higher-order elemental interactions.[24] This framework also included both internal cross-validation and external validation using independent datasets from disparate and geographically distinct cohorts. Central to our approach was the use of a layered, multi-tiered model stacking strategy, in which candidate models had to meet a pre-specified statistical performance threshold to advance to the subsequent stage of stacking. Moreover, to further mitigate the risk of model overfitting and spurious findings, we conducted detailed sensitivity analyses.

Overall, we observed slightly lower performance metrics in males compared to females. This discrepancy may reflect underlying differences in the phenotypic expression of autism between sexes. Some research suggests that autistic females, compared to their male counterparts, might require presenting with more autistic behaviors to be diagnosed.[70,71] It is also likely that given some autism phenotypes in females differ from those in males, such a discrepancy leads to underdiagnosis and particularly later diagnoses associated with masking of challenges, internalization of symptoms, and biases in diagnostic practices. Younger females with autism are often more motivated to fit in and may learn or mimic socially acceptable behaviors, masking their difficulties and making it harder to recognize their autism.[47]

There are several strengths of this work worth noting. Foremost, we have demonstrated that our biochemical test can rule out autism in children under 6 months, setting a new standard for autism screening. We collected hair samples from participants across a wide age range, allowing us to test the stability of our biochemical test across different developmental periods. We have validated our assay across multiple geographic regions, against varying degrees of outcome misclassification, mimicking plausible real-world clinical settings, and demonstrating its generalizability across different environmental exposures, diets, socio-economic factors, ethnicities, and other variables. The large sample size of our study allowed us to undertake a sensitivity analysis by age groups that align with the current recommended schedule of health checkups for children. This provides evidence for including this test as a routine tool for autism screening during children's regular visits to their pediatrician. Collection of hair samples is easy and non-invasive, requiring only a small amount of hair, making this diagnostic test readily implementable across various settings.

This study and the proposed diagnostic aid also have limitations that need to be addressed. Firstly, and most importantly, the diagnostic aid is designed as a rule-out tool and cannot deliver a positive diagnosis of autism. Its rule-in performance was not optimized, and thus it is not adequate for diagnosing autism. In this respect, it is similar to other tests primarily designed to rule out a health condition rather than confirm it (e.g., fecal DNA tests for colorectal cancer or multi-cancer blood screening assays).[72,73] A key focus of future directions, therefore, will be to develop a test that can be applied to the "non-negatives" for ruling in autism. This will enable us to assess individuals along a continuum of likelihood, a continuum of positive and negative diagnosis likelihood, allowing clinicians to triage referrals and organize service workflow more effectively. Moreover, the rule-out diagnostic aid could be combined with telehealth services for autism diagnosis, particularly for older children with non-negative diagnostic aid test results, to

yield a more balanced outcome.[74] Second, some subgroup sizes, particularly among female cases, were modest, which limits the precision of subgroup-specific estimates. The imbalance in sex distribution reflects the historically higher diagnosis rates of autism in males.[75] Future work involving larger datasets and newer techniques may offer greater sensitivity to weak signal patterns and warrant further exploration. Third, while the study spans multiple countries and diverse populations, capturing variation in ancestry, socioeconomic background, environmental exposures, and healthcare access, additional sites would enhance the generalizability of our findings.[76] Additional efforts will be required to identify biomarkers of phenotypes and comorbidities that enable individualized therapies to be delivered at an earlier age.

In conclusion, this diagnostic test may help clinicians identify children with low autism likelihood and who may not require immediate, resource-intensive autism evaluations, while ensuring that high-likelihood cases are not lost.

Funding Information: National Institute of Environmental Health Science (R35ES030435)

Competing Interest Declaration: Manish Arora is a founder and CEO of Linus Biotechnology Inc., a start-up company of the Mount Sinai Health System that develops hair-based biomarkers. He owns equity in the company and is listed as an inventor on patent applications, including hair biomarkers of ALS. Ghalib Bello, Louis A. Gomez, Sujeewa C. Piyankarage, Suzy Elhlou, Jyoti Chumber, Juliet Jaramilo, and Sophie Dessalle are employees of Linus Biotechnology Inc. Vishal Midya and Manuel Ruiz Marin advise on hair biomarker-related work at Linus Biotechnology Inc. All other authors declare no other competing financial interests or personal relationships that might influence the work reported in this paper.

Ethical Approval Statement: This current study is approved by the Mount Sinai IRB. The CHARGE and MARBLES study protocols were approved by the University of California at Davis and the State of California Institutional Review Boards. For SEAVER and PRISM, the present study was approved by the Mount Sinai IRB, and written informed consent was obtained from all participants or their parents or guardians. The RATSS study was approved by the Swedish Regional Ethical Review Board. The protocol for JECS was reviewed and approved by the IRB of the Japanese Ministry of the Environment's on Epidemiological Studies and by the Ethics Committees of all participating institutions. All participating women provided written informed consent. The study in Mexico City was approved by the ethical review board of Iberoamericana University. For all study populations included in this manuscript, informed consent was obtained after the nature and possible consequences of the studies were explained.

References

1. Zeidan J, Fombonne E, Scorah J, et al. Global prevalence of autism: A systematic review update. *Autism research* 2022; **15**(5): 778–90.
2. Shaw KA. Prevalence and early identification of autism spectrum disorder among children aged 4 and 8 years—Autism and Developmental Disabilities Monitoring Network, 16 Sites, United States, 2022. *MMWR Surveillance Summaries* 2025; **74**.
3. Nasir AK, Strong-Bak W, Bernard M. Diagnostic evaluation of autism spectrum disorder in pediatric primary care. *Journal of primary care & community health* 2024; **15**: 21501319241247997.

4. Gordon-Lipkin E, Foster J, Peacock G. Whittling down the wait time: exploring models to minimize the delay from initial concern to diagnosis and treatment of autism spectrum disorder. *Pediatric Clinics of North America* 2016; **63**(5): 851.
5. Bölte S, Girdler S, Marschik PB. The contribution of environmental exposure to the etiology of autism spectrum disorder. *Cellular and molecular life sciences* 2019; **76**(7): 1275–97.
6. Havdahl A, Niarchou M, Starnawska A, Uddin M, van der Merwe C, Warrier V. Genetic contributions to autism spectrum disorder. *Psychological medicine* 2021; **51**(13): 2260–73.
7. Lyall K, Croen L, Daniels J, et al. The changing epidemiology of autism spectrum disorders. *Annual review of public health* 2017; **38**(1): 81–102.
8. Modabbernia A, Arora M, Reichenberg A. Environmental exposure to metals, neurodevelopment, and psychosis. *Current opinion in pediatrics* 2016; **28**(2): 243–9.
9. Midya V, Alcala CS, Rechtman E, et al. Machine learning assisted discovery of interactions between pesticides, phthalates, phenols, and trace elements in child neurodevelopment. *Environmental Science & Technology* 2023; **57**(46): 18139–50.
10. Seabra CM, Aneichyk T, Erdin S, et al. Transcriptional consequences of MBD5 disruption in mouse brain and CRISPR-derived neurons. *Molecular autism* 2020; **11**(1): 45.
11. Curtin P, Austin C, Curtin A, et al. Dysregulated biodynamics in metabolic attractor systems precede the emergence of amyotrophic lateral sclerosis. *PLoS Computational Biology* 2020; **16**(4): e1007773.
12. Arora M, Giuliani A, Curtin P. Biodynamic interfaces are essential for human–environment interactions. *Bioessays* 2020; **42**(11): 2000017.
13. Huo M, Zhang J, Huang W, Wang Y. Interplay among metabolism, epigenetic modifications, and gene expression in cancer. *Frontiers in Cell and Developmental Biology* 2021; **9**: 793428.
14. Malavolta M, Piacenza F, Basso A, Giacconi R, Costarelli L, Mocchegiani E. Serum copper to zinc ratio: Relationship with aging and health status. *Mechanisms of ageing and development* 2015; **151**: 93–100.
15. Djoko KY, Cheryl-lynn YO, Walker MJ, McEwan AG. The role of copper and zinc toxicity in innate immune defense against bacterial pathogens. *Journal of Biological Chemistry* 2015; **290**(31): 18954–61.
16. Billman GE. Homeostasis: the underappreciated and far too often ignored central organizing principle of physiology. *Frontiers in physiology* 2020; **11**: 200.
17. Arora M, Austin C, Sarrafpour B, et al. Determining prenatal, early childhood and cumulative long-term lead exposure using micro-spatial deciduous dentine levels. *PloS one* 2014; **9**(5): e97805.
18. Austin C, Smith TM, Bradman A, et al. Barium distributions in teeth reveal early-life dietary transitions in primates. *Nature* 2013; **498**(7453): 216–9.
19. Cheng N, Rho JM, Masino SA. Metabolic dysfunction underlying autism spectrum disorder and potential treatment approaches. *Frontiers in molecular neuroscience* 2017; **10**: 34.
20. Frye RE. Mitochondrial dysfunction in autism spectrum disorder: unique abnormalities and targeted treatments. Seminars in pediatric neurology; 2020: Elsevier; 2020. p. 100829.
21. Wu Q, Ren Q, Meng J, Gao W-J, Chang Y-Z. Brain iron homeostasis and mental disorders. *Antioxidants* 2023; **12**(11): 1997.
22. Jeyasingh PD, Goos JM, Thompson SK, Godwin CM, Cotner JB. Ecological stoichiometry beyond redfield: an ionomic perspective on elemental homeostasis. *Frontiers in Microbiology* 2017; **8**: 722.
23. Austin C, Curtin P, Curtin A, et al. Dynamical properties of elemental metabolism distinguish attention deficit hyperactivity disorder from autism spectrum disorder. *Translational psychiatry* 2019; **9**(1): 238.
24. Midya V, Bello G, Andrew AS, Re DB, Stommel EW, Arora M. Dysregulation of hair-strand-based elemental biodynamics in amyotrophic lateral sclerosis. *Ebiomedicine* 2025; **119**.

25. Braddock B, Turner K, Kumarason K, et al. Improving Access to Autism Evaluation: A Coordinated Developmental-Behavioral Pediatrics and Pediatric Neurology Diagnostic Pathway. *Missouri Medicine* 2024; **121**(3): 225.
26. Brian JA, Zwaigenbaum L, Ip A. Standards of diagnostic assessment for autism spectrum disorder. *Paediatrics & child health* 2019; **24**(7): 444–51.
27. Hertz-Picciotto I, Croen LA, Hansen R, Jones CR, Van de Water J, Pessah IN. The CHARGE study: an epidemiologic investigation of genetic and environmental factors contributing to autism. *Environmental health perspectives* 2006; **114**(7): 1119–25.
28. Hertz-Picciotto I, Schmidt RJ, Walker CK, et al. A prospective study of environmental exposures and early biomarkers in autism spectrum disorder: design, protocols, and preliminary data from the MARBLES study. *Environmental health perspectives* 2018; **126**(11): 117004.
29. Berument SK, Rutter M, Lord C, Pickles A, Bailey A. Autism screening questionnaire: diagnostic validity. *The British Journal of Psychiatry* 1999; **175**(5): 444–51.
30. Mullen EM. Mullen scales of early learning: AGS Circle Pines, MN; 1995.
31. Siper PM, Kolevzon A, Wang AT, Buxbaum JD, Tavassoli T. A clinician-administered observation and corresponding caregiver interview capturing DSM-5 sensory reactivity symptoms in children with ASD. *Autism Research* 2017; **10**(6): 1133–40.
32. Colicino E, Cowell W, Pedretti NF, et al. Maternal steroids during pregnancy and their associations with exposure to lifetime stressful life events, prenatal stress appraisal and psychological functioning. *Psychoneuroendocrinology* 2023; **158**: 106395.
33. Baron-Cohen S, Hoekstra RA, Knickmeyer R, Wheelwright S. The autism-spectrum quotient (AQ)—adolescent version. *Journal of autism and developmental disorders* 2006; **36**(3): 343–50.
34. Bölte S, Willfors C, Berggren S, et al. The roots of autism and ADHD twin study in Sweden (RATSS). *Twin Research and Human Genetics* 2014; **17**(3): 164–76.
35. Ronald A, Larsson H, Anckarsäter H, Lichtenstein P. Symptoms of autism and ADHD: a Swedish twin study examining their overlap. *Journal of abnormal psychology* 2014; **123**(2): 440.
36. Srivastava AK, Schwartz CE. Intellectual disability and autism spectrum disorders: causal genes and molecular mechanisms. *Neuroscience & Biobehavioral Reviews* 2014; **46**: 161–74.
37. Thurm A, Farmer C, Salzman E, Lord C, Bishop S. State of the field: Differentiating intellectual disability from autism spectrum disorder. *Frontiers in psychiatry* 2019; **10**: 526.
38. Michikawa T, Nitta H, Nakayama SF, et al. Baseline profile of participants in the Japan Environment and Children's Study (JECS). *Journal of Epidemiology* 2018; **28**(2): 99–104.
39. Rajabi P, Noori AS, Sargolzaei J. Autism spectrum disorder and various mechanisms behind it. *Pharmacology Biochemistry and Behavior* 2024; **245**: 173887.
40. Mottron L, Bzdok D. Autism spectrum heterogeneity: fact or artifact? *Molecular psychiatry* 2020; **25**(12): 3178–85.
41. Coco MI, Dale R. Cross-recurrence quantification analysis of categorical and continuous time series: an R package. *Frontiers in psychology* 2014; **5**: 510.
42. Cofré R, Destexhe A. Entropy and Complexity Tools Across Scales in Neuroscience: A Review. *Entropy* 2025; **27**(2): 115.
43. Schreiber T. Measuring information transfer. *Physical review letters* 2000; **85**(2): 461.
44. Landa RJ, Reetzke R, Holingue CB, Herman D, Hess CR. Diagnostic stability and phenotypic differences among school-age children diagnosed with ASD before age 2. *Frontiers in Psychiatry* 2022; **13**: 805686.
45. Ozonoff S, Young GS, Belding A, et al. The broader autism phenotype in infancy: when does it emerge? *Journal of the American Academy of Child & Adolescent Psychiatry* 2014; **53**(4): 398–407. e2.
46. Girault JB, Piven J. The neurodevelopment of autism from infancy through toddlerhood. *Neuroimaging Clinics of North America* 2019; **30**(1): 97.

47. Bölte S, Neufeld J, Marschik PB, Williams ZJ, Gallagher L, Lai M-C. Sex and gender in neurodevelopmental conditions. *Nature Reviews Neurology* 2023; **19**(3): 136–59.
48. Wolpert DH. Stacked generalization. *Neural networks* 1992; **5**(2): 241–59.
49. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Machine learning* 2006; **63**(1): 3–42.
50. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Advances in neural information processing systems* 2017; **30**.
51. Ozonoff S, Young GS, Carter A, et al. Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. *Pediatrics* 2011; **128**(3): e488–e95.
52. Constantino JN, Zhang Y, Frazier T, Abbacchi AM, Law P. Sibling recurrence and the genetic epidemiology of autism. *American Journal of Psychiatry* 2010; **167**(11): 1349–56.
53. De Wit M, Morgan MJ, Libedinsky I, et al. A Systematic Review and Meta-Analysis: Research Using the Autism Polygenic Score. *medRxiv* 2024: 2024.03. 08.24303918.
54. Sun W, Poschmann J, Del Rosario RC-H, et al. Histone acetylome-wide association study of autism spectrum disorder. *Cell* 2016; **167**(5): 1385–97. e11.
55. Jensen AR, Lane AL, Werner BA, McLees SE, Fletcher TS, Frye RE. Modern biomarkers for autism spectrum disorder: future directions. *Molecular diagnosis & therapy* 2022; **26**(5): 483–95.
56. Saghazadeh A, Ataeinia B, Keynejad K, Abdolalizadeh A, Hirbod-Mobarakeh A, Rezaei N. A meta-analysis of pro-inflammatory cytokines in autism spectrum disorders: Effects of age, gender, and latitude. *Journal of psychiatric research* 2019; **115**: 90–102.
57. Frye RE, Slattery JC, Quadros EV. Folate metabolism abnormalities in autism: potential biomarkers. *Biomarkers in medicine* 2017; **11**(8): 687–99.
58. Lingampelly SS, Naviaux JC, Heuer LS, et al. Metabolic network analysis of pre-ASD newborns and 5-year-old children with autism spectrum disorder. *Communications Biology* 2024; **7**(1): 536.
59. Rose S, Niyazov DM, Rossignol DA, Goldenthal M, Kahler SG, Frye RE. Clinical and molecular characteristics of mitochondrial dysfunction in autism spectrum disorder. *Molecular diagnosis & therapy* 2018; **22**(5): 571–93.
60. Joseph RM, Fricker Z, Fenoglio A, Lindgren KA, Knaus TA, Tager-Flusberg H. Structural asymmetries of language-related gray and white matter and their relationship to language function in young children with ASD. *Brain imaging and behavior* 2014; **8**(1): 60–72.
61. Li D, Karnath H-O, Xu X. Candidate biomarkers in children with autism spectrum disorder: a review of MRI studies. *Neuroscience bulletin* 2017; **33**(2): 219–37.
62. Jones W, Klin A. Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature* 2013; **504**(7480): 427–31.
63. Mason L, Shic F, Falck-Ytter T, et al. Preference for biological motion is reduced in ASD: implications for clinical trials and the search for biomarkers. *Molecular autism* 2021; **12**(1): 74.
64. Zamzow RM, Ferguson BJ, Ragsdale AS, Lewis ML, Beversdorf DQ. Effects of acute beta-adrenergic antagonism on verbal problem solving in autism spectrum disorder and exploration of treatment response markers. *Journal of clinical and experimental neuropsychology* 2017; **39**(6): 596–606.
65. Curtin P, Austin C, Curtin A, et al. Dynamical features in fetal and postnatal zinc-copper metabolic cycles predict the emergence of autism spectrum disorder. *Science advances* 2018; **4**(5): eaat1293.
66. Arora M, Reichenberg A, Willfors C, et al. Fetal and postnatal metal dysregulation in autism. *Nature communications* 2017; **8**(1): 15493.
67. Ij H. Statistics versus machine learning. *Nat Methods* 2018; **15**(4): 233.
68. Austin C, Curtin P, Arora M, et al. Elemental dynamics in hair accurately predict future autism spectrum disorder diagnosis: an international multi-center study. *Journal of Clinical Medicine* 2022; **11**(23): 7154.

69. Shmueli G. To explain or to predict? *Statistical science* 2010: 289–310.
70. Lai M-C, Lombardo MV, Auyeung B, Chakrabarti B, Baron-Cohen S. Sex/gender differences and autism: setting the scene for future research. *Journal of the American Academy of Child & Adolescent Psychiatry* 2015; **54**(1): 11–24.
71. Corbett BA, Schwartzman JM, Libsack EJ, et al. Camouflaging in autism: Examining sex-based and compensatory models in social cognition and communication. *Autism Research* 2021; **14**(1): 127–42.
72. Neal RD, Johnson P, Clarke CA, et al. Cell-free DNA–based multi-cancer early detection test in an asymptomatic screening population (NHS-Galleri): design of a pragmatic, prospective randomised controlled trial. *Cancers* 2022; **14**(19): 4818.
73. Fisher DA, Princic N, Miller-Wilson L-A, Wilson K, Fendrick AM, Limburg P. Utilization of a colorectal cancer screening test among individuals with average risk. *JAMA network open* 2021; **4**(9): e2122269–e.
74. Sutherland R, Trembath D, Roberts J. Telehealth and autism: A systematic search and review of the literature. *International journal of speech-language pathology* 2018; **20**(3): 324–36.
75. Posserud MB, Skretting Solberg B, Engeland A, Haavik J, Klungsøyr K. Male to female ratios in autism spectrum disorders by age, intellectual disability and attention-deficit/hyperactivity disorder. *Acta Psychiatrica Scandinavica* 2021; **144**(6): 635–46.
76. Cheney AM, Barrera T, Rodriguez K, Jaramillo López AM. The intersection of workplace and environmental exposure on health in latinx farm working communities in rural Inland Southern California. *International journal of environmental research and public health* 2022; **19**(19): 12940.