Machine-learning analysis of temporal molecular dynamics stratifies autism likelihood - a multinational study

Vishal Midya, PhD, MSTAT

Assistant Professor

Department of Environmental Medicine

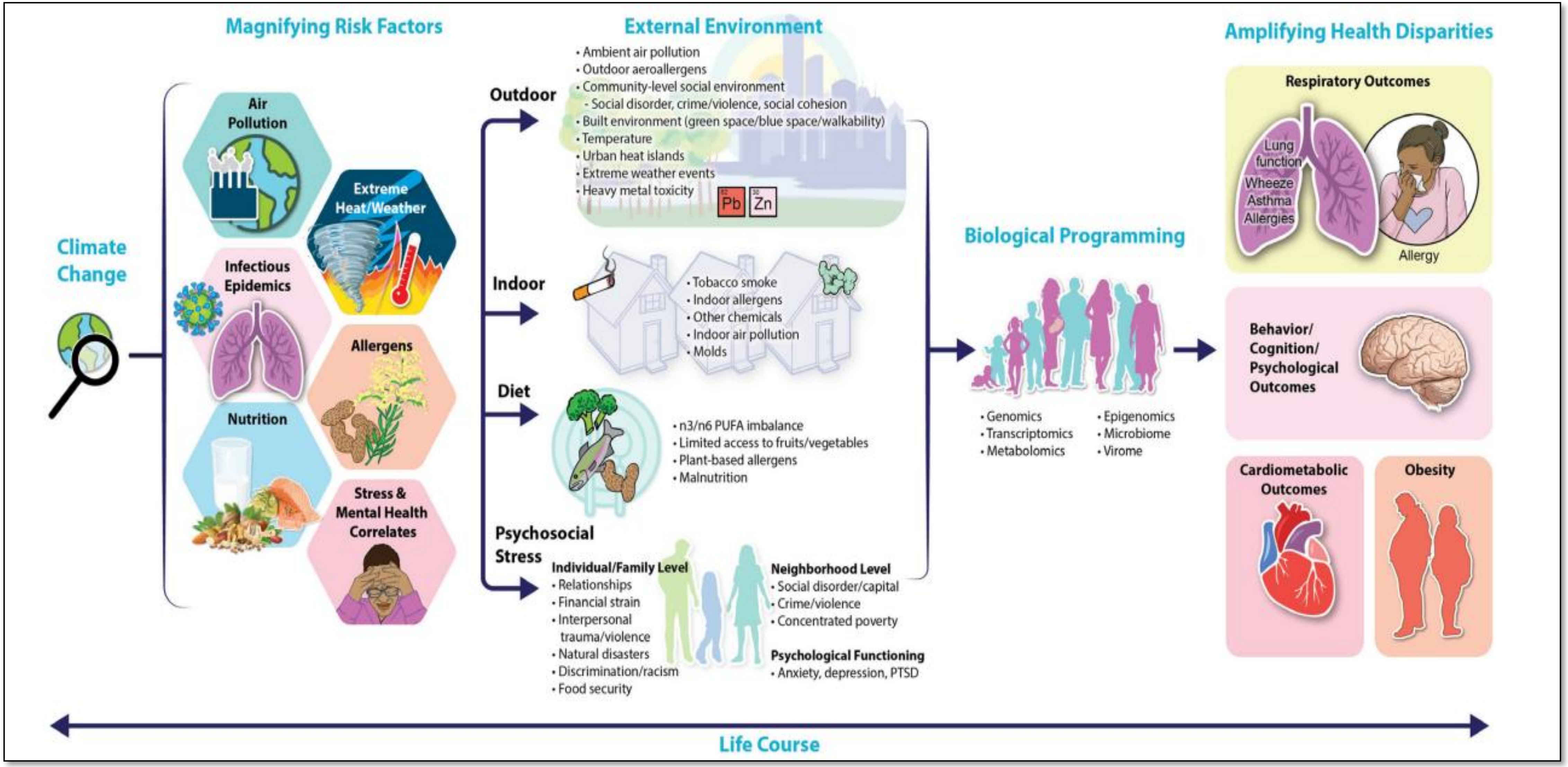
Department of Public Health

Icahn School of Medicine, Mount Sinai, NY, USA

Icahn School of Medicine at Mount Sinai

AI × Statistics: From Intelligent Models to Making Real-World Differences

What is Exposome ?



Supplement From Science Magazine and the Icahn School of Medicine at Mount Sinai (https://www.science.org/content/resource/frontiers-medical-research)

Why Study the Exposome?

1) Understanding causes of disease

"…70-90% of disease risks are probably due to differences in environments"

EPIDEMIOLOGY

Environment and Disease Risks

Stephen M. Rappaport and Martyn T. Smith

A new paradigm is needed to assess how a lifetime of exposure to environmental factors affects the risk of developing chronic diseases.

Although the risks of developing chronic diseases are attributed to both genetic and environmental factors, 70 to 90% of disease risks are probably due to differences in environments (1–3). Yet, epidemiologists increasingly use genome-wide association studies (GWAS) to investigate diseases, while relying on questionnaires to characterize "environmental exposures." This is because GWAS represent the only approach for exploring the totality of any risk factor (genes, in this case) associated with disease prevalence. Moreover, the value of costly genetic information is diminished when inaccurate and imprecise environmental data lead to biased inferences regarding gene-environment interactions (4). A more comprehensive and quantitative view of environmental exposure is needed if epidemiologists are to discover the major causes of chronic diseases.

An obstacle to identifying the most important environmental exposures is the fragmentation of epidemiological research along lines defined by different factors. When epidemiologists investigate environmental risks, they tend to concentrate on a particular category of exposures involving air and water pollution, occupation, diet and obesity, stress and behavior, or types of infection. This slicing of the disease pie along parochial lines leads to scientific separation and confuses the definition of "environmental exposures." In fact, all of these exposure categories can contribute to chronic diseases and should be investigated collectively rather than separately.

To develop a more cohesive view of environmental exposure, it is important to recognize that toxic effects are mediated through chemicals that alter critical molecules, cells, and physiological processes inside the body. Thus, it would be reasonable to consider the "environment" as the body's internal chemical environment and "exposures" as the amounts of biologically active chemicals in this internal environment. Under this view, exposures are not restricted to chemicals (toxicants) entering the body from air, water, or food, for example, but also include chemicals produced by inflammation, oxidative stress, lipid peroxidation, infections, gut flora, and other natural processes (5, 6) (see the figure). This internal chemical environment continually fluctuates during life due to changes in external and internal sources, aging, infections, life-style, stress, psychosocial factors, and preexisting diseases.

The term "exposome" refers to the totality of environmental exposures from conception onwards, and has been proposed to be a

School of Public Health, University of California, Berkeley, CA 94720–7356, USA. E-mail: srappaport@berkeley.edu

460 22 OCTOBER 2010 VOL 330 SCIENCE www.sciencemag.org

Published by AAAS

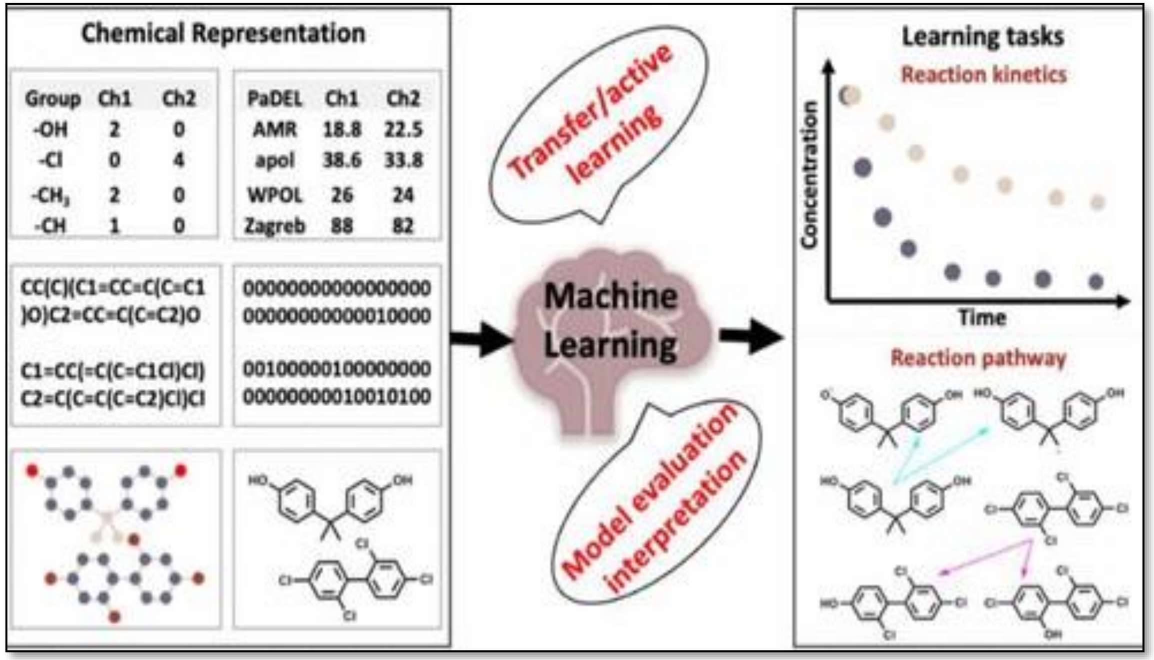
2) Ensuring chemical safety and human/ecological health

Content from J. Sobus

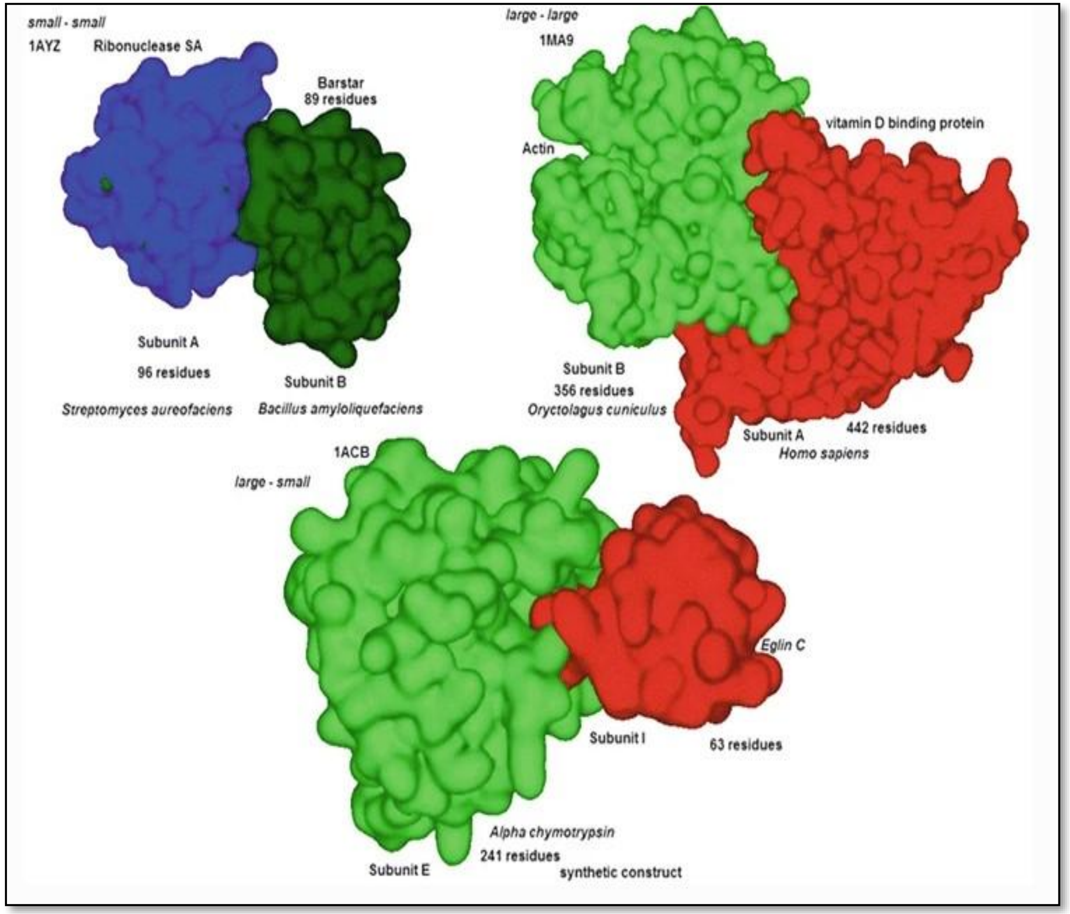
It’s all about Interactions

The concept of "interaction" has been construed in many ways through different scientific fields.

- **Chemistry:**



- **Protein-Protein Interaction:**



1) Kai Zhang and Huichun Zhang, ACS ES&T Water
2) Kangueane, P., Nilofer,C. (2018). Principles of Protein-Protein Interaction.

Dynamical metal-metal recurrence interaction

ScienceAdvances

Current Issue First release papers Archive About

HOME > SCIENCE ADVANCES > VOL. 4, NO. 5 > DYNAMICAL FEATURES IN FETAL AND POSTNATAL ZINC-COPPER METABOLIC CYCLES PREDICT THE...

RESEARCH ARTICLE | DISEASES AND DISORDERS

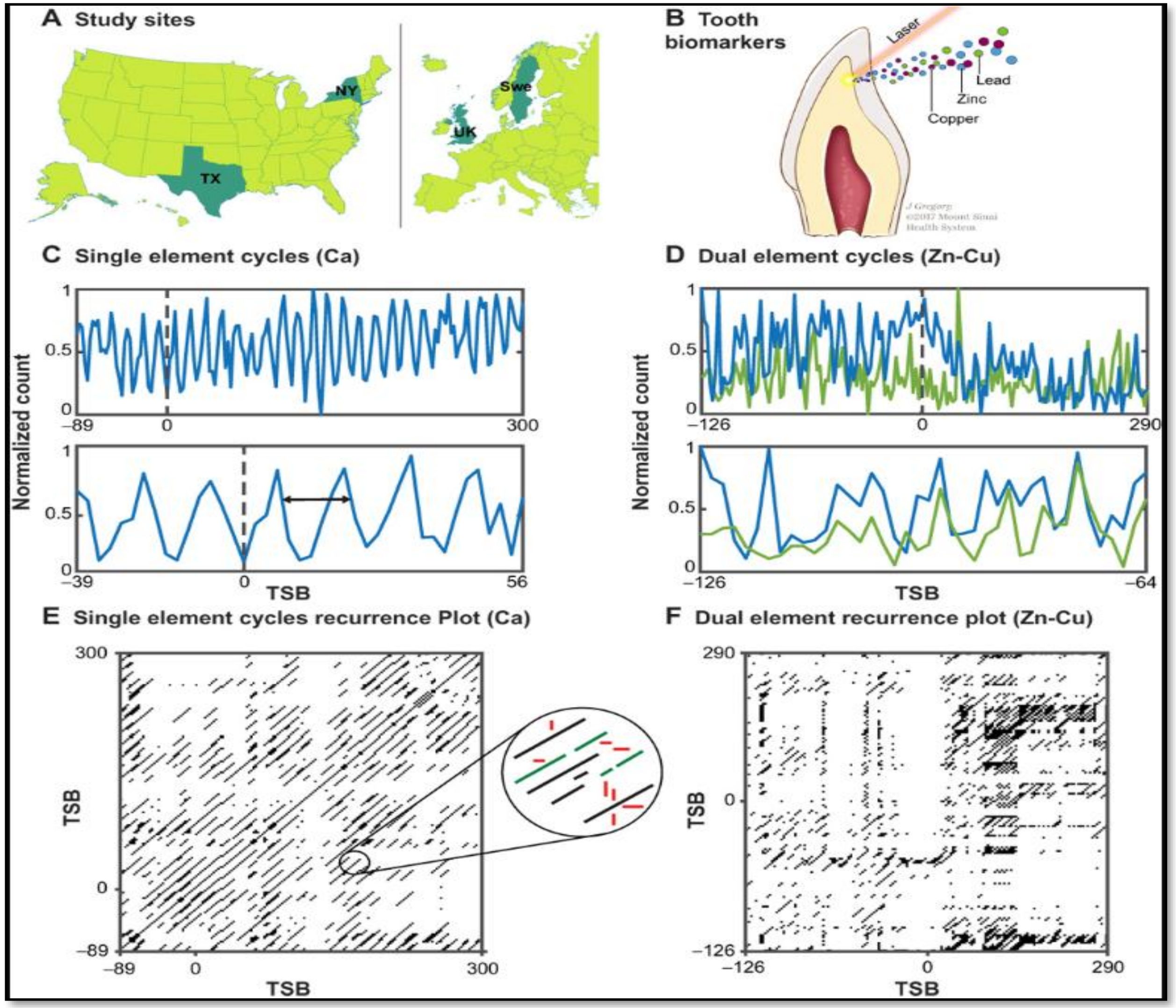
Dynamical features in fetal and postnatal zinc-copper metabolic cycles predict the emergence of autism spectrum disorder

PAUL CURTIN, CHRISTINE AUSTIN, AUSTEN CURTIN, CHRIS GENNINGS, MANISH ARORA, (FOR THE EMERGENT DYNAMICAL SYSTEMS GROUP), KRISTIINA TAMMIMIES, CHARLOTTE WILLFORS, STEVE BERGGREN, [...], AND ABRAHAM REICHENBERG +12 authors Authors Info & Affiliations

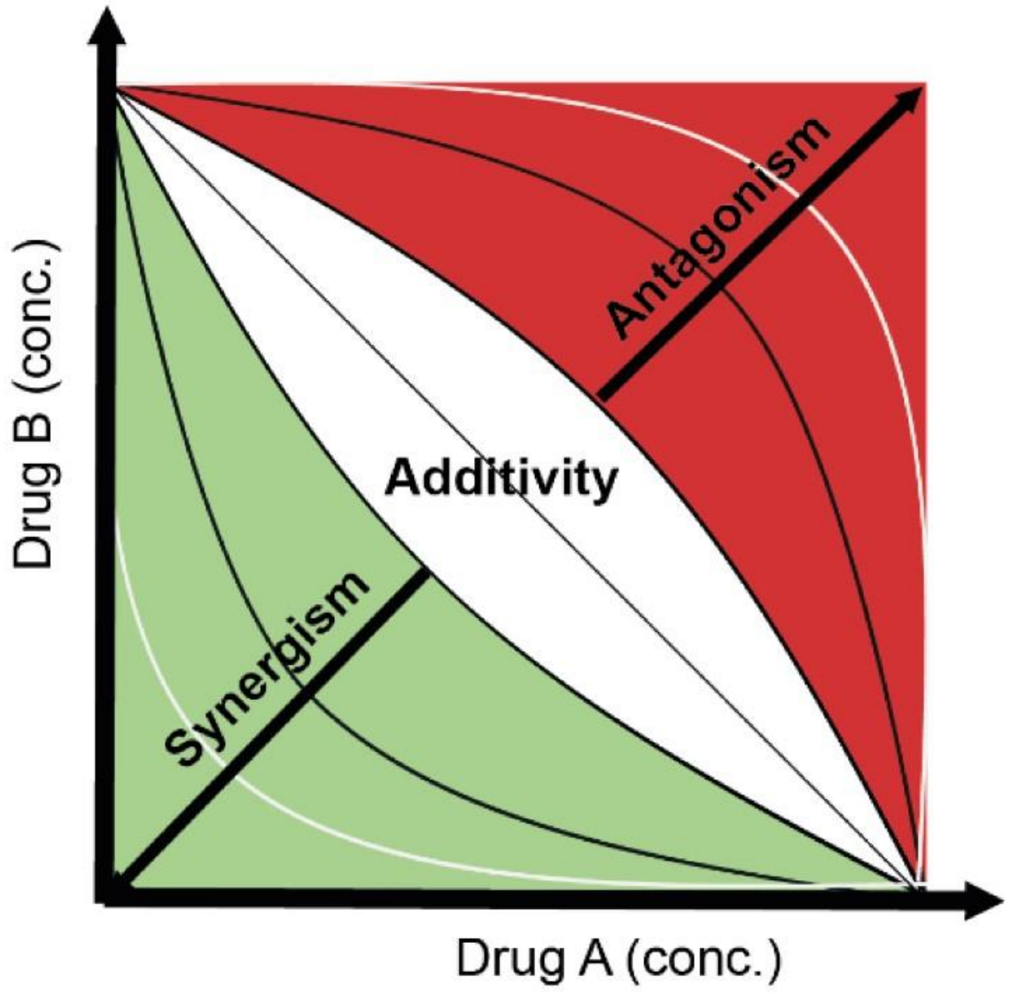
SCIENCE ADVANCES • 30 May 2018 • Vol 4, Issue 5 • DOI: 10.1126/sciadv.aat1293

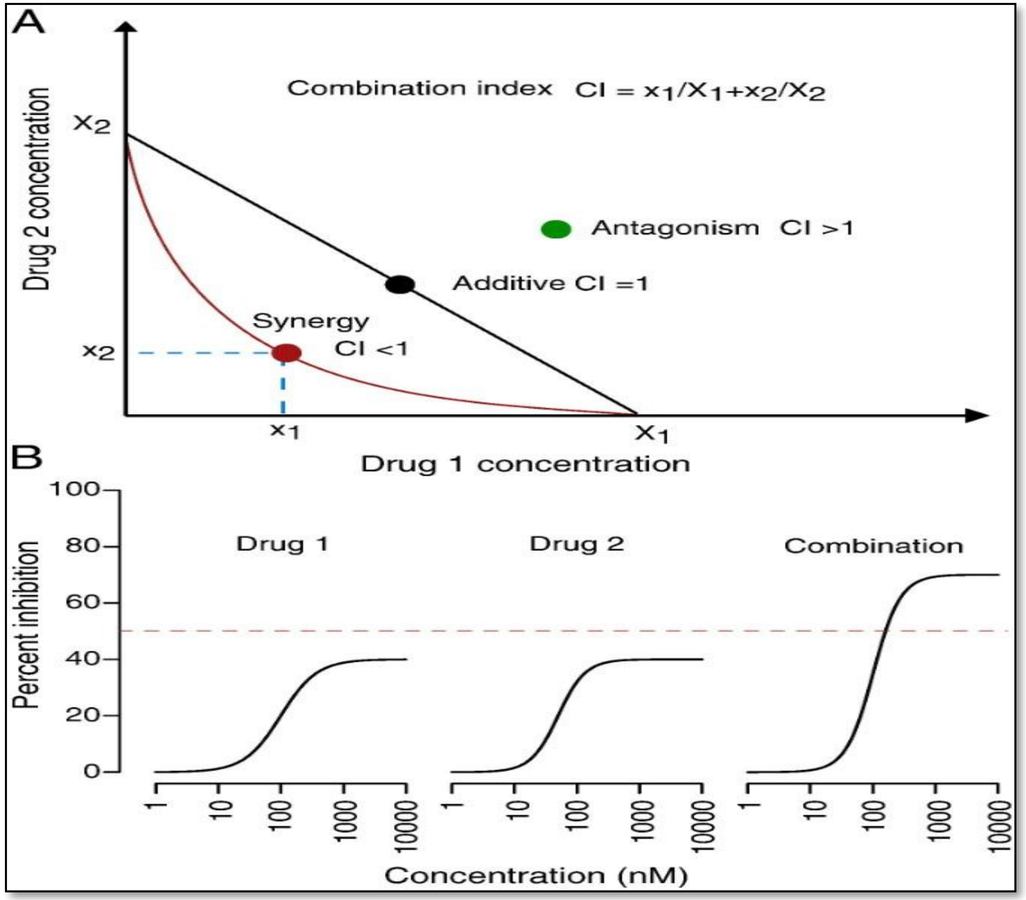
Abstract

Metals are critical to neurodevelopment, and dysregulation in early life has been documented in autism spectrum disorder (ASD). However, underlying mechanisms and biochemical assays to distinguish ASD cases from controls remain elusive. In a nationwide study of twins in Sweden, we tested whether zinc-copper cycles, which regulate metal metabolism, are disrupted in ASD. Using novel tooth-matrix biomarkers that provide direct measures of fetal elemental uptake, we developed a predictive model to distinguish participants who would be diagnosed with ASD in childhood from those who did not develop the disorder. We replicated our findings in three independent studies in the United States and the UK. We show that three quantifiable characteristics of fetal and postnatal zinc-copper rhythmicity are altered in ASD: the average duration of zinc-copper cycles, regularity with which the cycles recur, and the number of complex features within a cycle. In all independent study sets and in the pooled analysis, zinc-copper rhythmicity was disrupted in



Toxicological Interaction





- Interactions are individualized
- Interactions only happen when the concentrations are beyond or below some thresholds

1) Rodea-Palomares, Ismael, Miguel González-Pleiter, Keila Martín-Betancor, Roberto Rosal, and Francisca Fernández-Piñas. 2015.
2) Bhagwan Yadav, Krister Wennerberg, Tero Aittokallio, Jing Tang, 2015

Something out of Nothing ?

Something from "Nothing" – Eight Weak Estrogenic Chemicals Combined at Concentrations below NOECs Produce Significant Mixture Effects

Elisabete Silva, Nissanka Rajapakse, and Andreas Kortenkamp

View Author Information

Cite this: *Environ. Sci. Technol.* 2002, 36, 8, 1751–1756
Publication Date: March 14, 2002
https://doi.org/10.1021/es0101227

Environmental Toxicology and Chemistry

Environmental Toxicology

Mixture toxicity and interactions of copper, nickel, cadmium, and zinc to barley at low effect levels: Something from nothing?

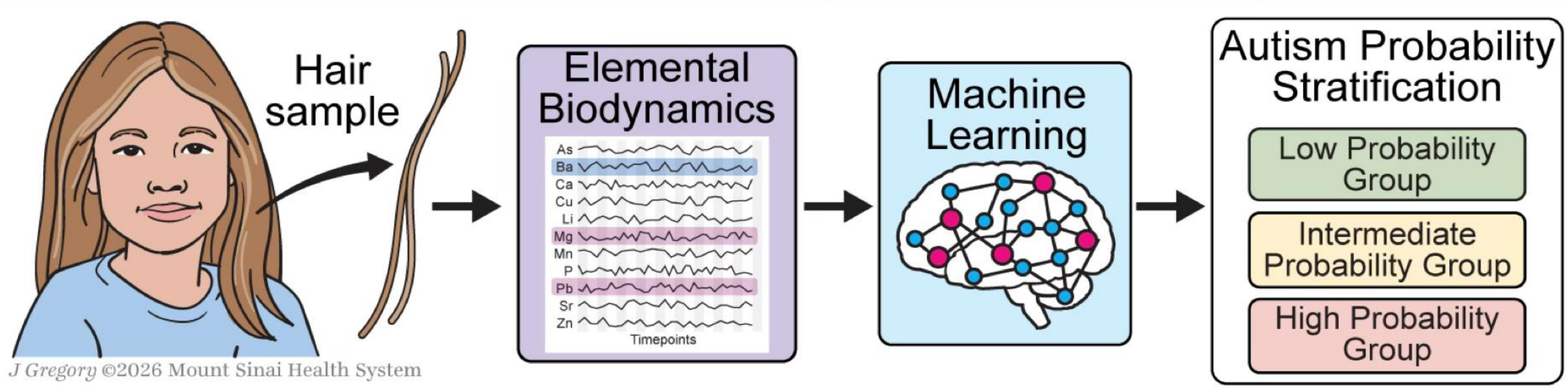
Liske Versieren, Steffie Evers, Karel De Schamphelaere, Ronny Blust, Erik Smolders

First published: 23 January 2016 | https://doi.org/10.1002/etc.3380 | Citations: 29

The need for an autism likelihood stratification model

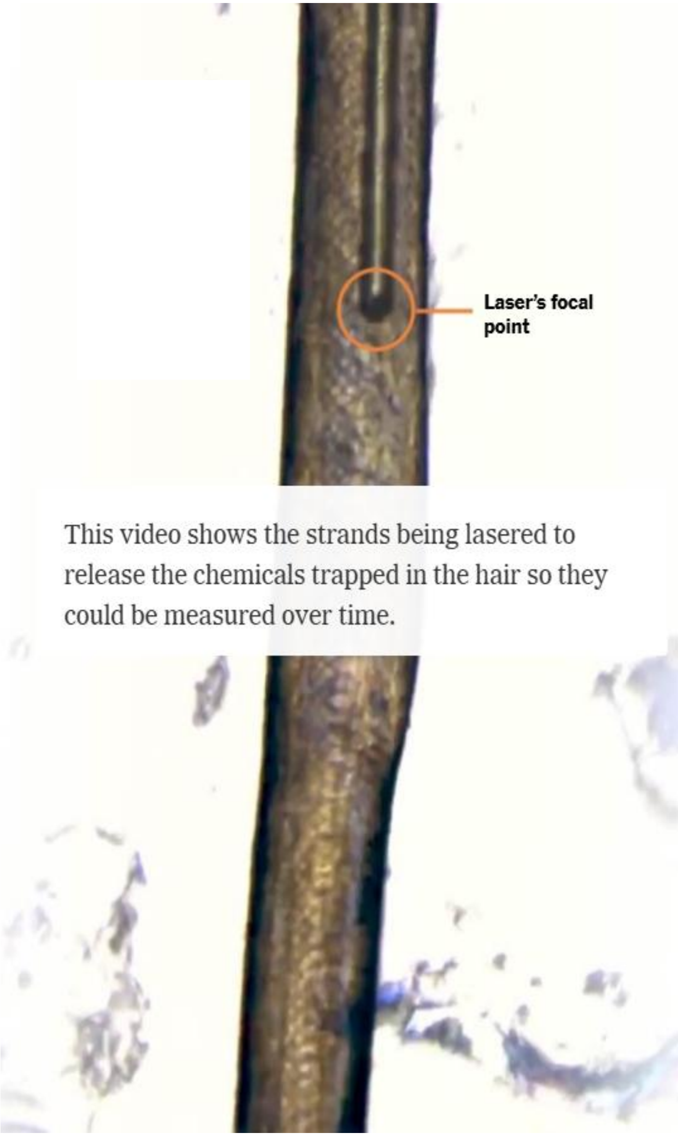
- Autism is now diagnosed in approximately 1 to 3% of all children in most mid to high-income countries.
- Due to increasing referral rates and persistent shortages in clinicians specializing in autism diagnosis, wait times for autism diagnostic assessment may be months to years.
- As a result, the median age of diagnosis in the US and other nations remains over 4 years, hindering timely intervention.
- When considering children referred from general paediatricians for autism evaluation by specialists, a significant proportion of children do not ultimately get diagnosed with autism, but all children undergo lengthy behavioural assessments that place demands on clinician time.

The need for an autism likelihood stratification model



Using noninvasively collected hair strands, we devise and describe a novel machine-learning-driven two-stage triage approach that leverages the time-varying biodynamics of elemental intensities to stratify the likelihood of autism.

Hair-based Elemental Intensity Data



The New York Times

How Did This Family End Up Back in a Toxic House?

https://www.nytimes.com/interactive/2025/12/29/us/insurers-smoke-damaged-homes-toxins.html

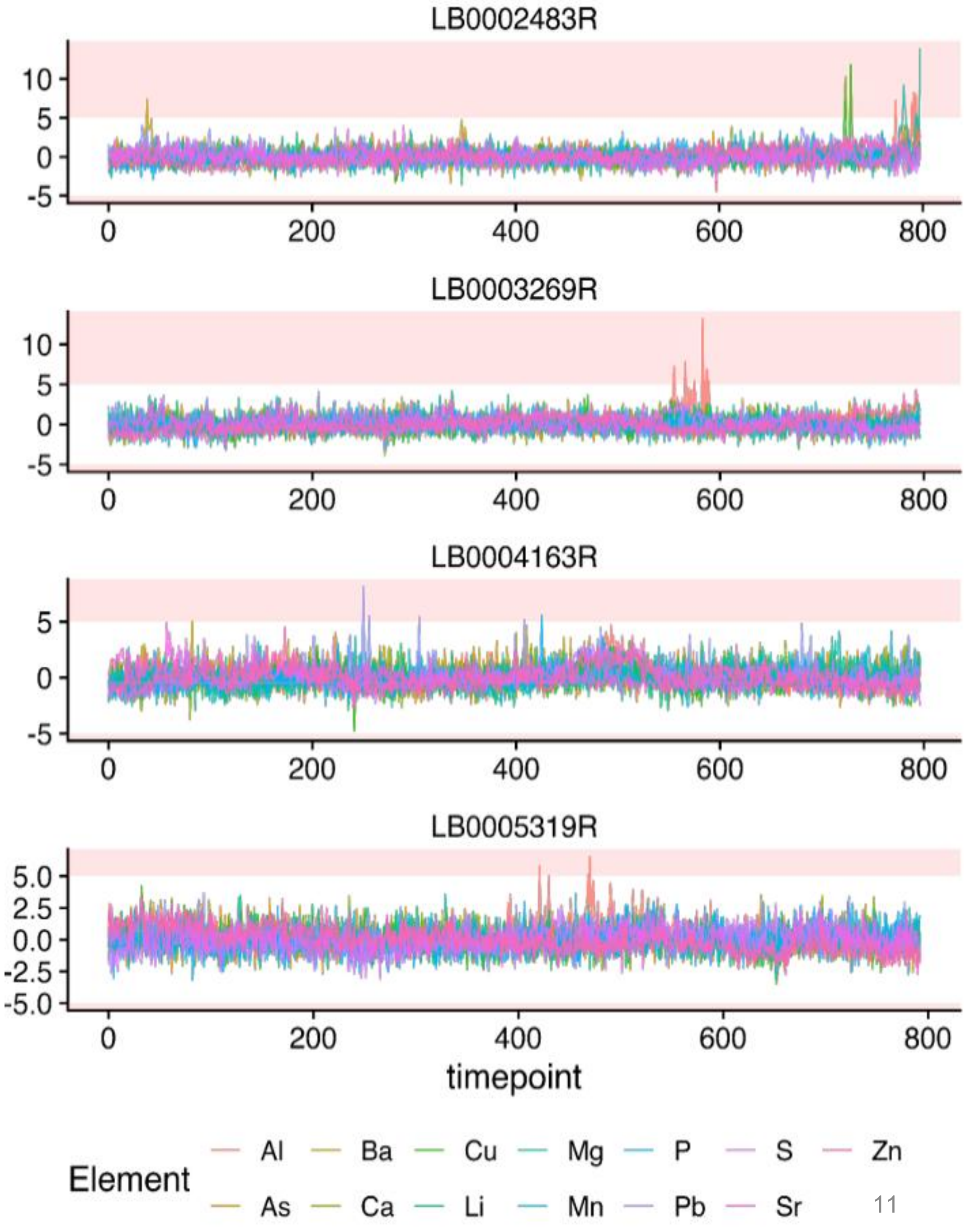
Hair-based Elemental Intensity Data

- Single hair strands from each participant were washed to remove surface contaminants in a solution of 1% Triton X-100 and ultrapure water using sonication for 1 minute. Hairs were then rinsed with ultrapure water to remove the surfactant and dried in an oven at 60 °C overnight.

- Elemental analysis of the hair strands was performed using a laser ablation (LA) system equipped with a 193 nm excimer laser coupled to an inductively coupled plasma-mass spectrometer (8900 triple quadrupole ICP-MS)

- 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. Downstream feature extraction and modeling were carried out using 12 elements: magnesium (Mg), phosphorus (P), calcium (Ca), copper (Cu), zinc (Zn), strontium (Sr), barium (Ba), lead (Pb), lithium (Li), manganese (Mn), Arsenic (As), and Sulphur (S).



Overall schematic of cohorts used in the study

- **CHARGE (California, US, n=825, Autism=164, non-autism=661)**
  Female, n(%) : 189 (22.91%)

- **MARBLES (California, US, n=287, Autism=66, non-autism=221)**
  Female, n(%) : 127 (44.3%)

- **RATSS (Sweden, n=306, Autism=70, non-autism=236)**
  Female, n(%) : 136 (44.44%)

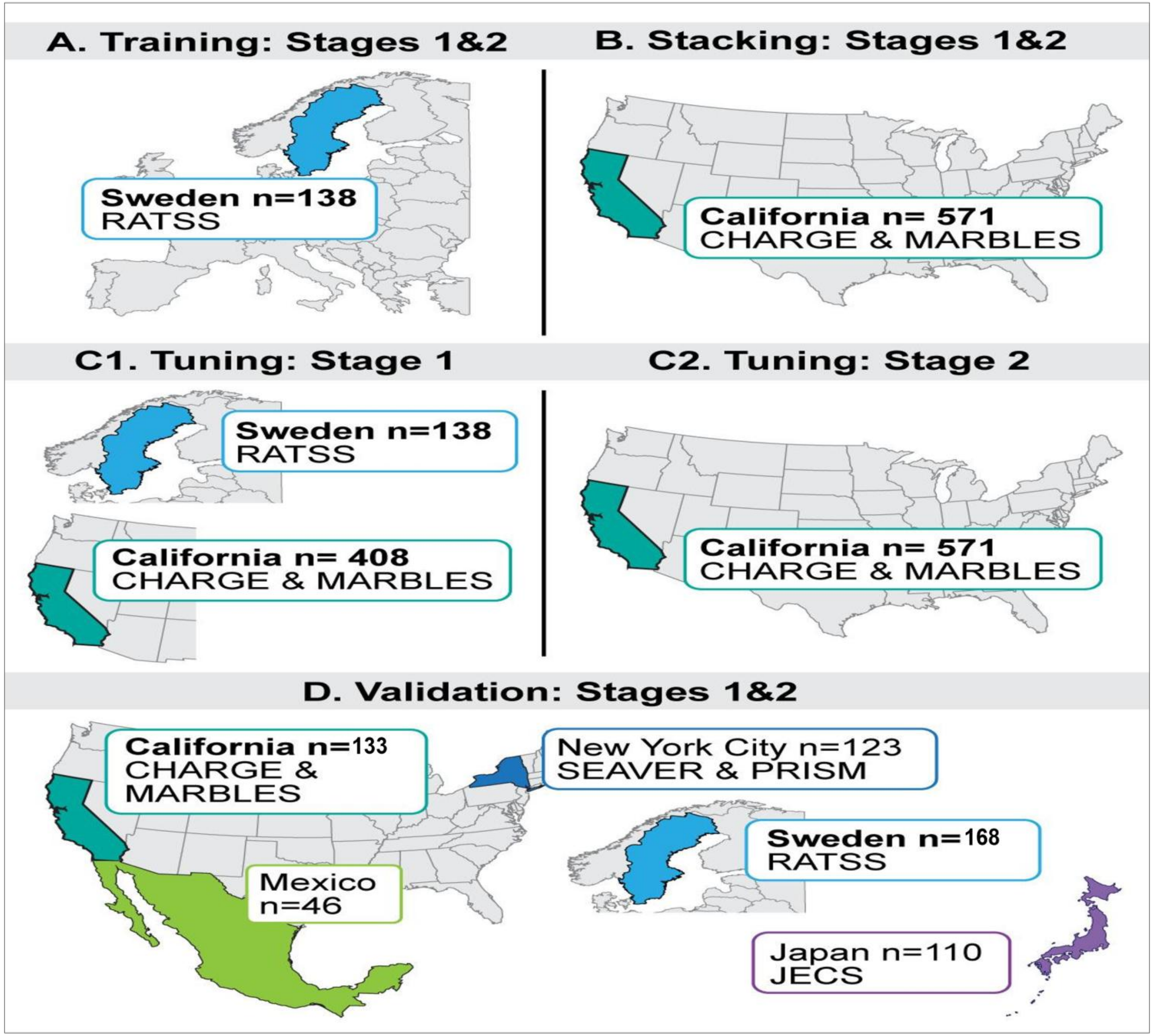
- **SEAVER (New York City, US, n=39, Autism=21, non-autism=18)**
  Female, n(%) : 14 (35.90%)

- **Mexico (Mexico City, n=46, Autism=33, non-autism=13)**
  Female, n(%) : 12 (26.09%)

- **PRISM (New York City, US, n=84, Autism=0, non-autism=84)**
  Female, n(%) : 33 (39.29%)

- **JECS (Japan, n=110, Autism=0, non-autism=110)**
  Female, n(%) : 55 (50%)

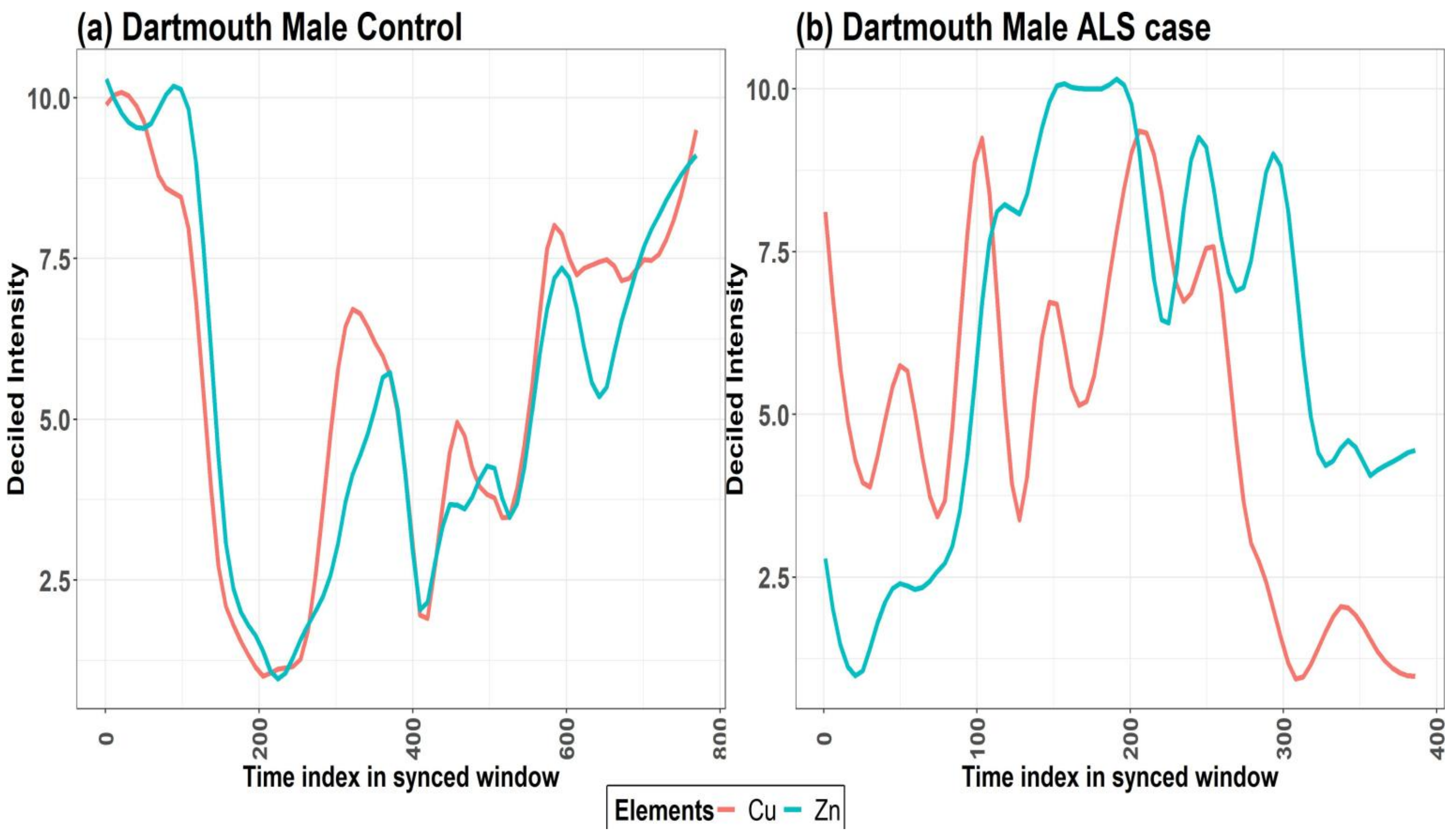
Overall schematic of cohorts used in the study



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

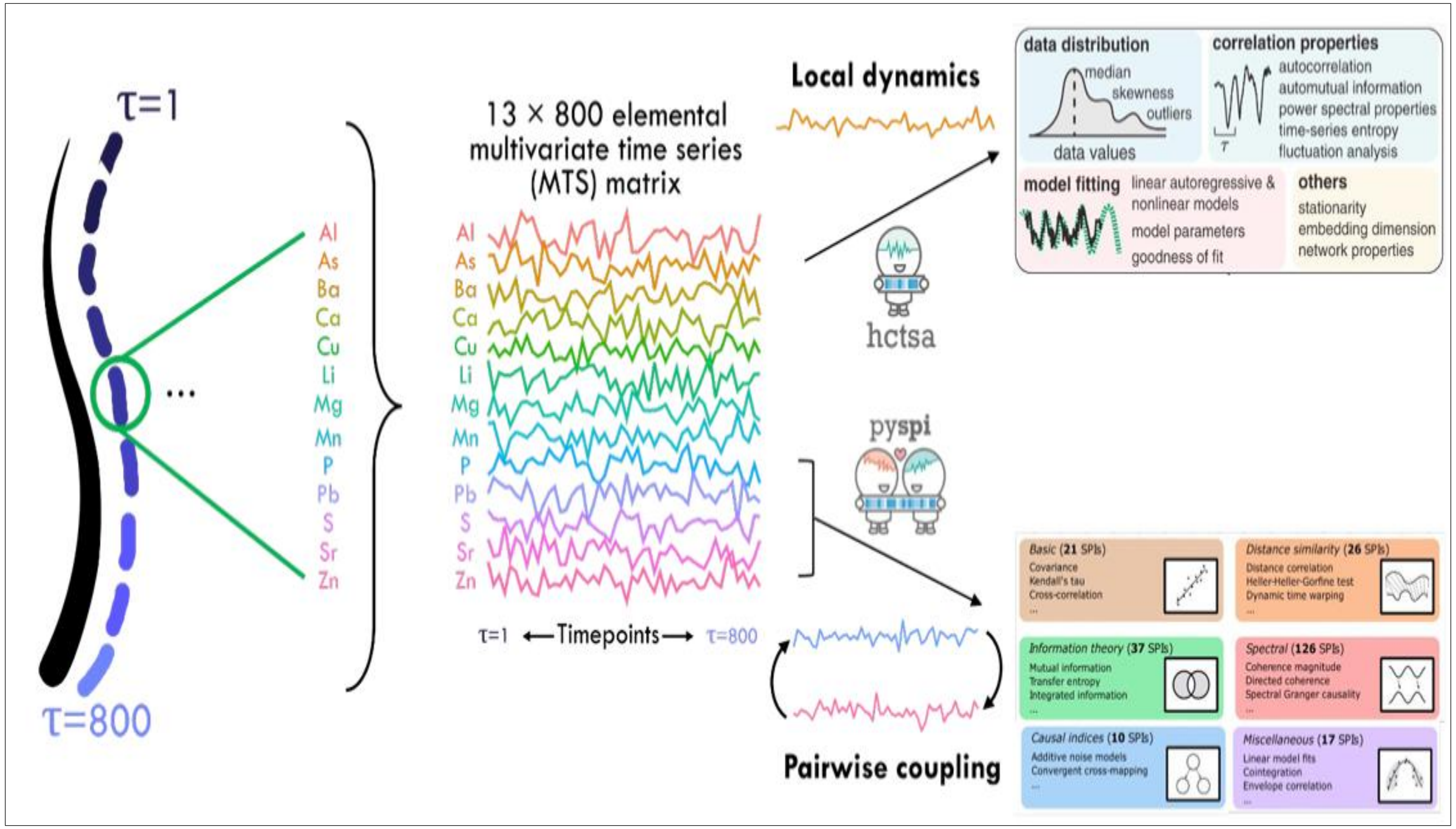
Elemental Biodynamics, a marker for dysregulation in homeostasis

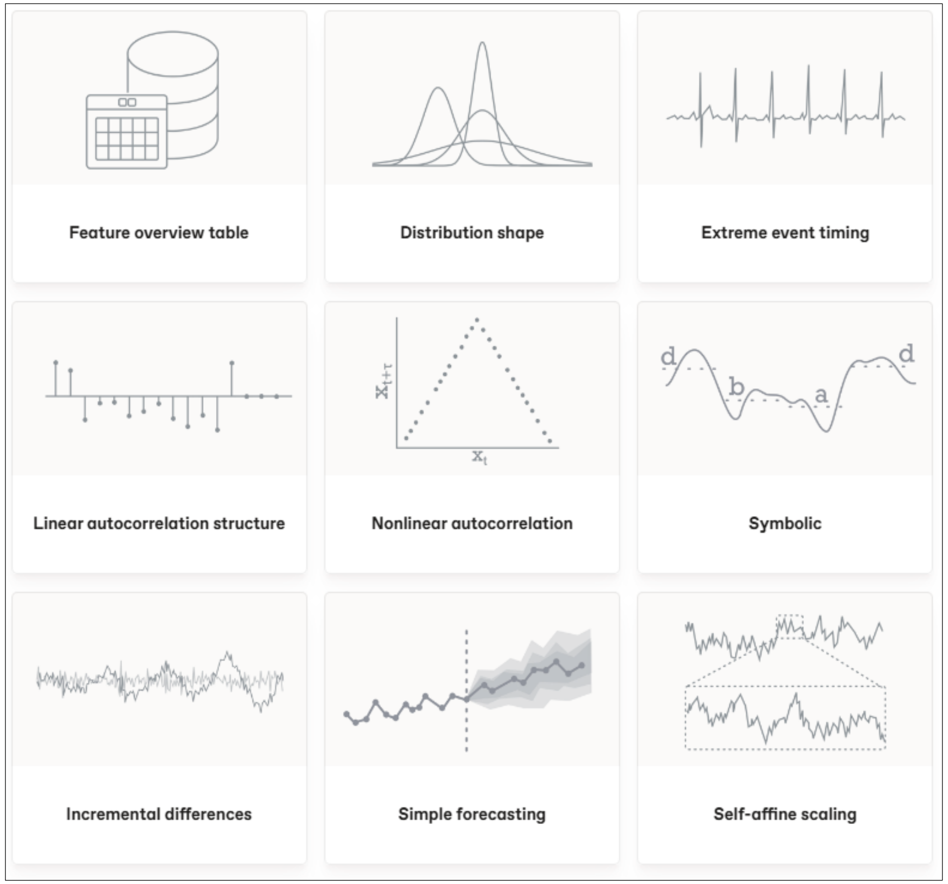
- Many biological processes (e.g., sleep patterns) have temporal biodynamic rhythms.
- Since elemental homeostasis is tightly controlled by stable regulatory mechanisms that respond to environmental signals, we hypothesize that the longitudinal dynamic patterns of elemental intensities (referred to as elemental biodynamics) may serve as a marker of homeostatic alterations.
- Delineating such measures of elemental homeostasis may serve as a predictor of disease/disorder.



Midya, V., Bello, G., Andrew, A. S., Re, D. B., Stommel, E. W., & Arora, M. (2025). Dysregulation of hair-strand-based elemental biodynamics in amyotrophic lateral sclerosis. *Lancet Ebiomedicine*, 119.

Features of Biodynamics





catch22: CAnonical Time-series CHaracteristics

Features of Biodynamics

- 8 feature extraction techniques to characterize various aspects of the temporality:

- Cross-recurrence quantification analysis (CRQA), which quantifies rhythmicity and synchronicity of elemental time series
- Entropy and complexity measures (Entropy-Complexity, Multiscale Cross-Entropy) that quantify irregularity and unpredictability in time series
- Causal network centrality (which assesses the structural importance of each element within a network inferred from inter-element relationships)
- Distributional features (Summary Statistics), and methods that quantify features of temporality (Catch22, TSFEL)

Illustration with Centrality-based time-series Network

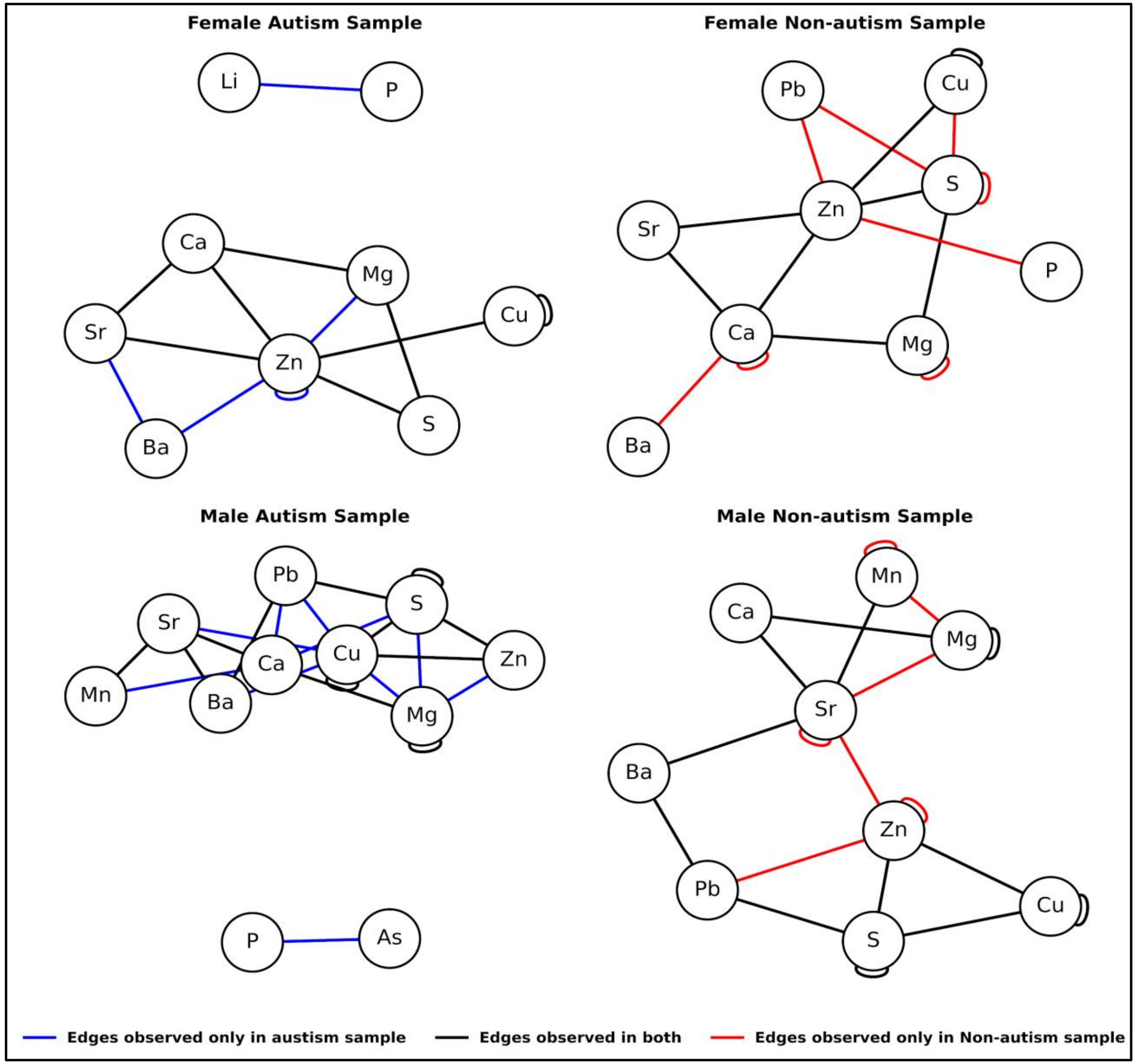
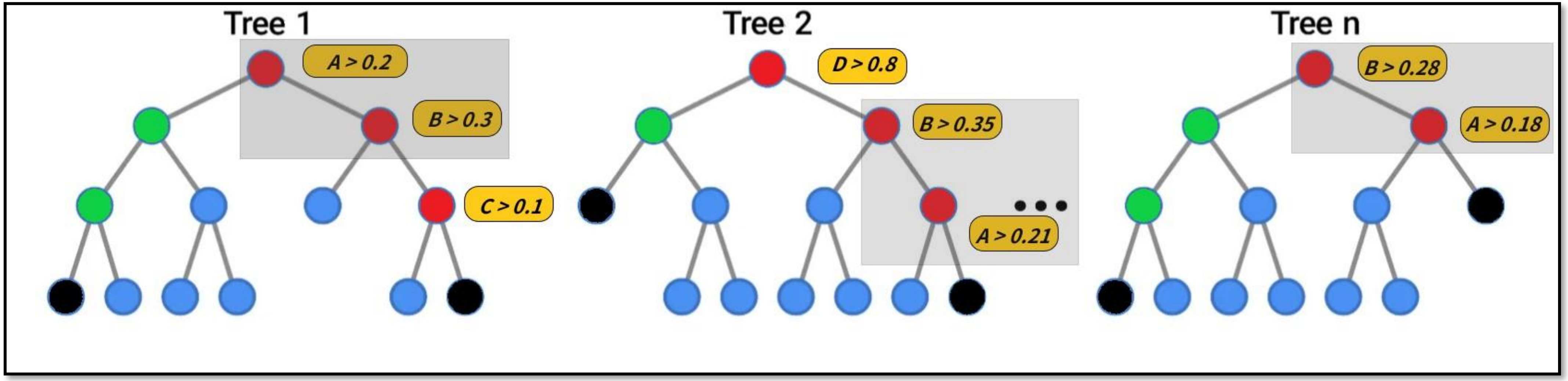


Illustration of interaction networks created using time-series network analysis between a sample with and without autism in our validation dataset, showing differences in connectivity among elements. Note that an edge between elements indicates a significant association (corrected for false discoveries at a significance level of 0.05) regardless of direction or strength. An edge in black indicates edges that exist in both the autism and non-autism samples; blue indicates edges that exist only in the autism sample; and red indicates edges that exist only in the non-autism sample.

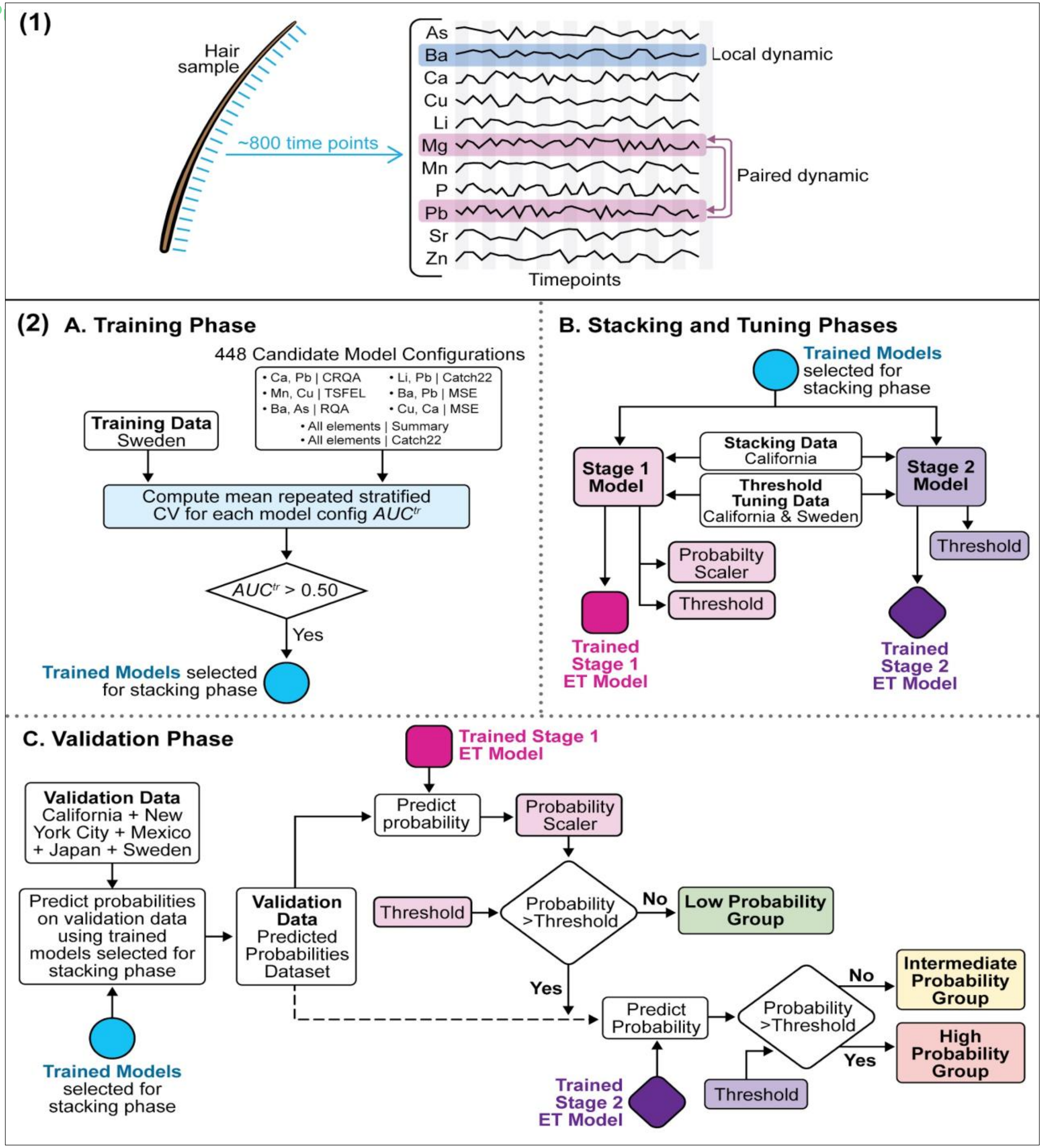
Need for tree-based models



Midya, V., Nagdeo, K., Lane, J. M., Torres-Olascoaga, L. A., Martínez, G. G., Horton, M. K., ... & Eggers, S. (2024). Akkermansia muciniphila attenuates association between specific metal exposures during pregnancy and depressive symptoms in late childhood. *iScience*, *27*(12).

Midya, Vishal, Cecilia Sara Alcala, …., Chris Gennings, Maria J. Rosa, and Damaskini Valvi. "Machine learning assisted discovery of interactions between pesticides, phthalates, phenols, and trace elements in child neurodevelopment." *Environmental Science & Technology* 57, no. 46 (2023): 18139-18150.

Midya, Vishal, and Chris Gennings. "Detecting Shape-Based Interactions Among Environmental Chemicals Using an Ensemble of Exposure-Mixture Regression and Interpretable Machine Learning Tools." *Statistics in Biosciences* (2023): 1-21.



Modelling Schema

The figure presents an overview of

(1) . The temporality of elemental intensities from a hair strand, and our modeling approach.

(2A) denotes the training phase where models were trained using data from the RATSS cohort (Sweden),

(2B) denotes the stacking phases in which the 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 cohorts; once the models are finalized and tested on multiple cohorts, threshold tuning is performed.

ET: Extra Trees Classifier. AUC^{tr}: training AUC

Results: Predictions

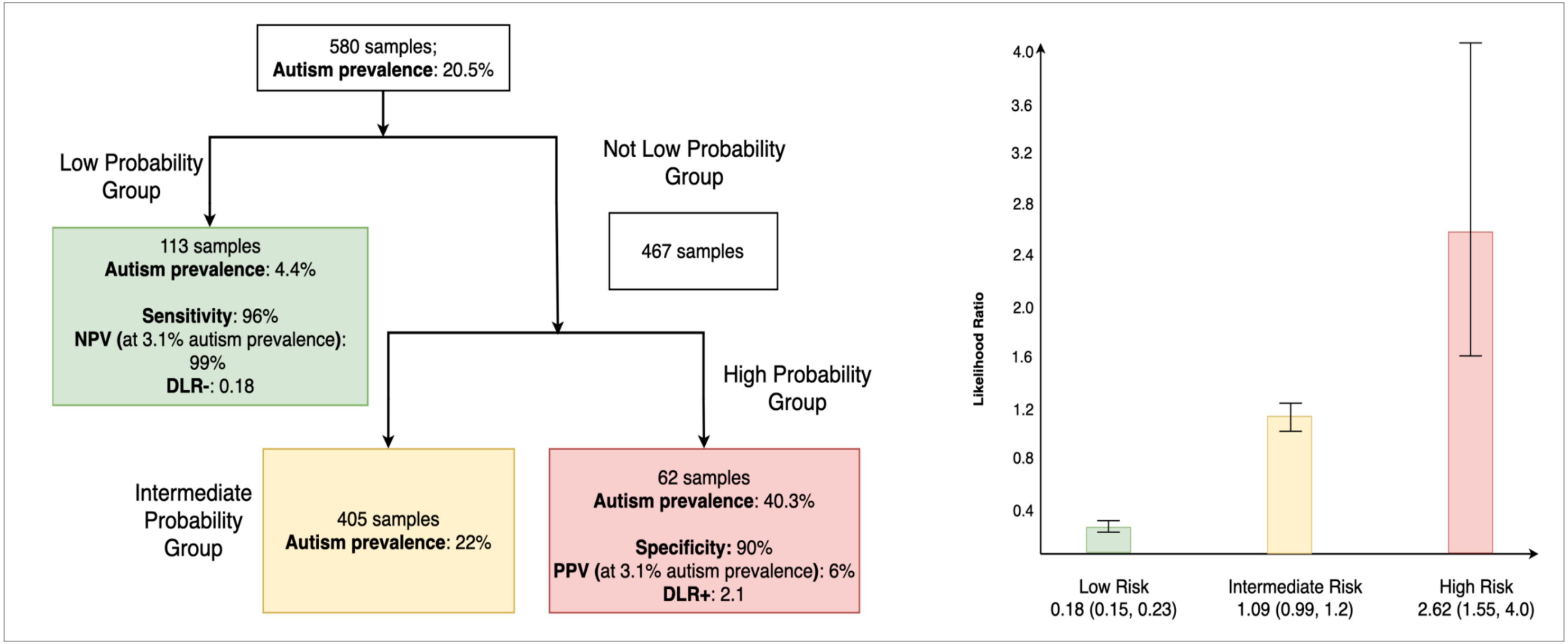


Table 1: Overall performance metrics for the likelihood stratification framework

Sex	Sample Size	No. of Autism	Odds Ratio	Overall Autism Prevalence (%)	Predicted Probability Group	n (%)	Autism Prevalence by Group (%)	Group-Specific Likelihood Ratio (95% CI)
Male	360	107	2.28 (1.43, 3.62)	29.72	Low	41 (11.39)	9.76	0.26 (0.21,0.32)
					Intermediate	264 (73.33)	29.92	1.01 (0.89,1.13)
					High	55 (15.28)	43.64	1.83 (1.12,3.04)
Female	220	12	3.55 (1.02, 12.34)	5.45	Low	72 (32.73)	1.39	0.24 (0.16,0.50)
					Intermediate	141 (64.09)	7.09	1.32 (0.98,1.52)
					High	7 (3.18)	14.29	2.89 (0.0,15.21)
Total	580	119	2.41 (1.56, 3.72)	20.52	Low	113 (19.48)	4.42	0.18 (0.15,0.22)
					Intermediate	405 (69.83)	21.98	1.09 (0.98,1.20)
					High	62 (10.69)	40.32	2.62 (1.63,4.02)

Results: Feature Interpretations

- SHAP (SHapley Additive exPlanations) analyses of both the Stage 1 and Stage 2 meta-learners on the stacking set indicated distributed contributions rather than reliance on any single base model.
- In females, the highest-weighted contributors were **Li-Zn** and **Li-Ba** biodynamics (based on elemental network centrality features).
- In males, contributions were particularly diffuse, with prominent pairs such as **P-Ba** (using multiscale cross-entropy features), **Mg-Cu**, and, particularly for the Stage 1 meta-learner, **Mg-Zn** (using Catch22 features).

This suggests that the discriminative signal for autism is distributed across multiple dynamical representations and multiple inter-element relationships.

Concluding Remarks

- We introduce a novel biochemical hair-based machine-learning tool designed to stratify autism probability in children as young as one month.

- In our proof-of-concept results from multiple populations, the likelihood stratification tool achieved a high sensitivity of 96% in stratifying non-autism controls as low probability, whereas the second-stage model achieved a specificity of 90% in identifying a high-probability group among participants not triaged as low probability.

- Because this biochemical stratification tool can be implemented in the first year of life, it can also reduce the age at initiation of therapy, which has been shown to have greater benefits when started earlier.

Public

Acknowledgement

Great Science requires a great team effort!

- **Icahn School of Medicine at Mount Sinai**: Jamil M. Lane, Cecilia S. Alcala, Manasi Agrawal, Chris Gennings, Maria Rosa, Dania Valvi, Rosalind Wright, Robert Wright, Manish Arora
- **University of Iowa**: Shoshannah Eggers
- **University of California at Davis**: Deborah H. Bennett, Rebecca J. Schmidt
- **Karolinska Institute**: Sven Bölte
- **IS Global**: Nuria Güil-Oumrait, Martine Vrijheid
- **INSP México**: Mara Téllez Rojo
- **LinusBio:** Ghalib A. Bello, Louis A. Gomez, Sujeewa C. Piyankarage

- **Funding**: National Institute of Environmental Health Sciences

Icahn School of Medicine at Mount Sinai — Institute for Exposomic Research

The University of Iowa

ISGlobal Barcelona Institute for Global Health

INSP

UCDAVIS HEALTH | MIND INSTITUTE

Karolinska Institutet

LINUSBIO

Thank You

Email: vishal.midya@mssm.edu

Icahn School of Medicine at Mount Sinai