

# Supplement for Elemental dynamics in hair accurately predict future autism spectrum disorder diagnosis: an international multi-center study

Christine Austin<sup>1,2†</sup>, Paul Curtin<sup>1,2†\*</sup>, Manish Arora<sup>1,2†\*</sup>, Abraham Reichenberg<sup>1,3†</sup>, Austen Curtin<sup>1,2†</sup>, Miyuki Iwai-Shimada<sup>4</sup>, Robert Wright<sup>1</sup>, Rosalind Wright<sup>1,5</sup>, Karl Lundin Remnelius<sup>6</sup>, Johan Isaksson<sup>6,9</sup>, Sven Bölte<sup>6,7,8</sup>, Shoji F. Nakayama<sup>4</sup>

<sup>1</sup> Linus Biotechnology Inc., New York, NY 10013, USA

<sup>2</sup> Environmental Medicine and Public Health, Mount Sinai School of Medicine, New York, NY 10029, USA

<sup>3</sup> Seaver Autism Center, Department of Psychiatry, Mount Sinai School of Medicine, New York, NY 10029, USA

<sup>4</sup> Exposure Dynamics Research Section, Health and Environmental Risk Division, National Institute for Environmental Studies, Tsukuba 305-8506, Japan

<sup>5</sup> Department of Pediatrics, Mount Sinai School of Medicine, New York, NY 10029, USA

<sup>6</sup> Center of Neurodevelopmental Disorders (KIND), Centre for Psychiatry Research, Department of Women's and Children's Health, Karolinska Institutet and Stockholm Health Care Services, Region Stockholm, 11330 Stockholm, Sweden

<sup>7</sup> Department of Medical Sciences, Child and Adolescent Psychiatry Unit, Uppsala University, 75185 Uppsala, Sweden

<sup>8</sup> Child and Adolescent Psychiatry, Stockholm Health Care Services, Region Stockholm, 11861 Stockholm, Sweden

<sup>9</sup> Curtin Autism Research Group, Curtin School of Allied Health, Curtin University, Perth, WA 6102, Australia

\* Correspondence: paul.curtin@linusbio.com (P.C.); manish.arora@mssm.edu (M.A.)

† These authors contributed equally to this work.

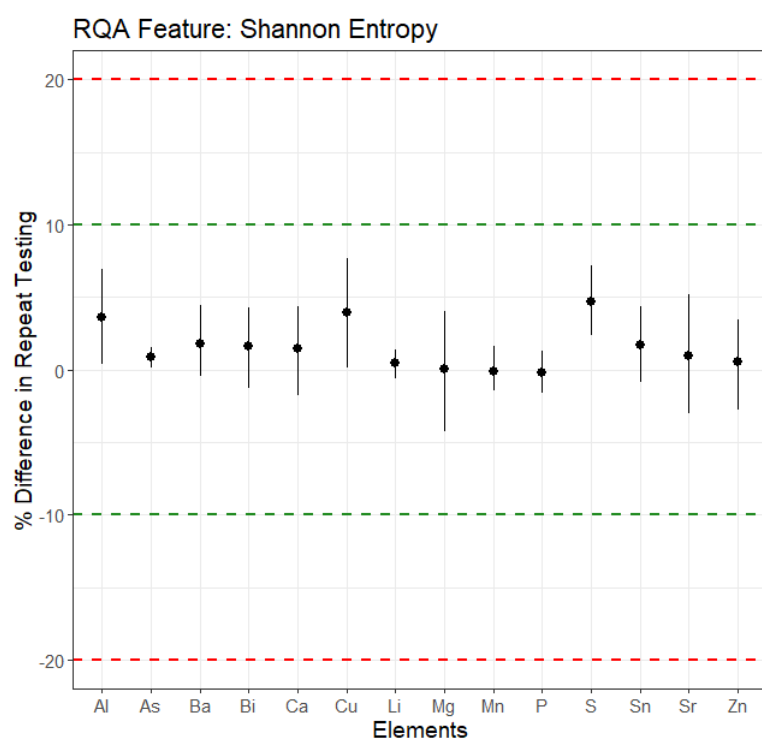
**Index**

1. Quality Assurance and Quality Control Procedures for Laboratory Hair Analysis
2. Derivation of Descriptive Features
3. Model Performance – Confusion Matrix
4. Participant Comorbidity Characteristics

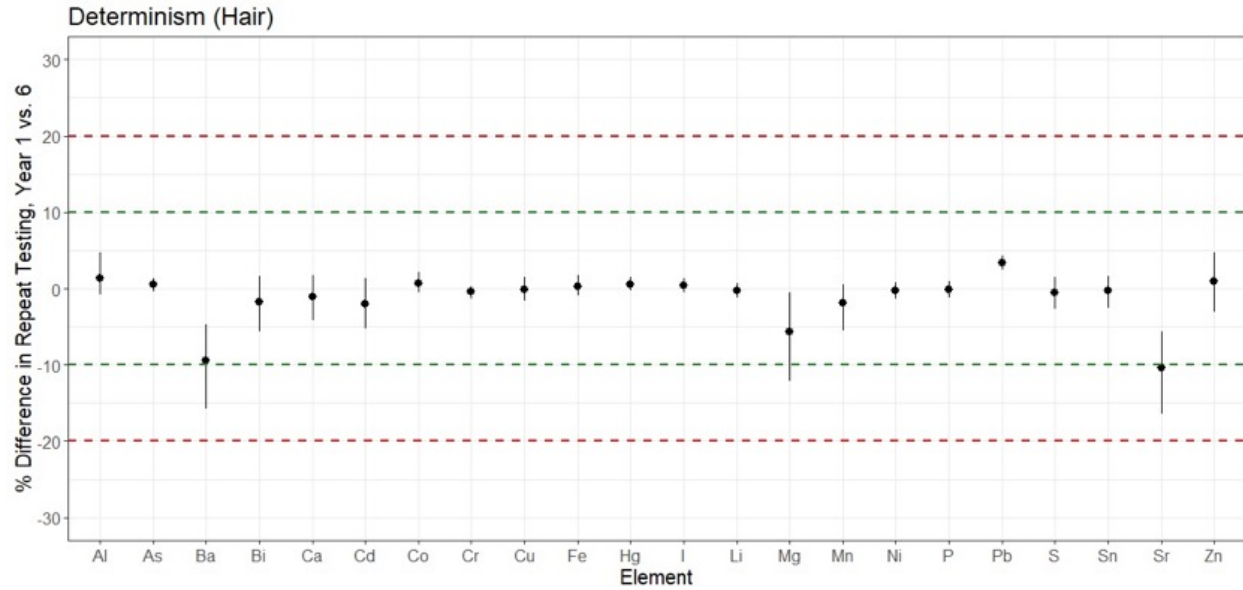
## I. QA/QC for Laboratory Analysis of Hair Samples

### Reproducibility of results

A second hair strand was analyzed for a subset of 20 samples for QA/QC. Element signal profiles from duplicates are aligned and the percent difference between each sequential data point measured and averaged. A difference of  $< \pm 20\%$  (recovery  $100 \pm 20\%$ ) was considered acceptable. Percent difference between duplicates is typically less than 10% for most elements (see Figure S1).



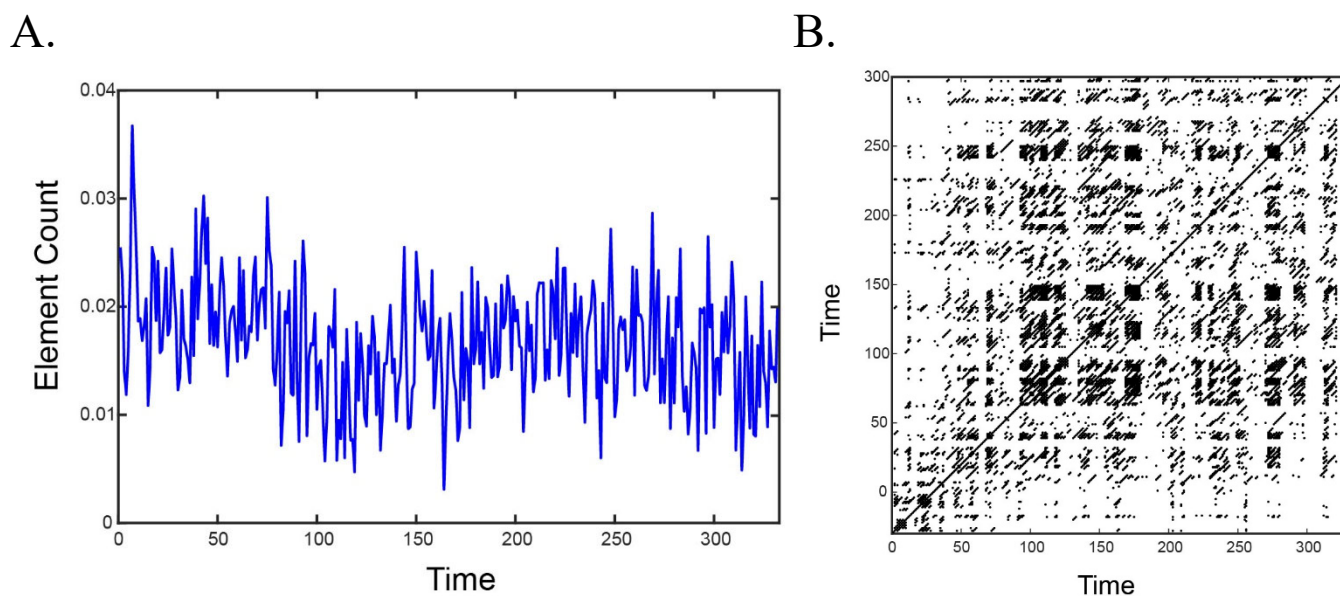
**Figure S1.** Average percent difference across elements in bioinformatic features – here, Shannon Entropy, estimated via Recurrence Quantification Analysis – measured in two hairs collected from the same individual at the same time but analyzed by LA-ICP-MS months apart. Green dashed line indicates  $\pm 10\%$  difference between duplicates; red dashed line indicates  $\pm 20\%$  difference between duplicates.  $n=20$  participants.



**Figure S2.** Average percent difference across elements in bioinformatic features – here, Determinism, estimated via Recurrence Quantification Analysis – measured in two hairs collected from the same individual 5 years apart (at age 1 year and then at age 6 years). Green dashed line indicates  $\pm 10\%$  difference between duplicates; red dashed line indicates  $\pm 20\%$  difference between duplicates.  $n=20$  participants.

## II. Derivation of Descriptive Features

For each of 15 elemental time series measured in each hair, and for the pairwise interactions between each elemental time series, a recurrence matrix or cross-recurrence matrix, respectively, was generated to reconstruct underlying signal dynamics. Supplementary Figure S3 provides an example of this process in the analysis of a single hair trace. The formation of diagonal lines in a recurrence matrix indicates the formation and dissolution of transient periodic processes. Likewise, the emergence of vertical and horizontal lines, referred to as laminar structures, indicates periods of relative signal stability. Recurrence quantification analysis (RQA) involves the quantitation of these features to characterize the magnitude and prevalence of these features in different signals; likewise, cross-recurrence quantification analysis (CRQA) extends this approach to consider pairwise dependencies between signals. From each recurrence and cross-recurrence matrix derived from each sample, quantitative metrics are derived, as described in accompanying Table S1.



**Figure S3. Recurrence analysis of elemental data derived from hair.** In left panel, (A), elemental (Phosphorus) time-series data extracted from analysis of a hair sample are shown. In right panel, (B), a recurrence matrix is shown. This formation of diagonal and vertical lines in the recurrence matrix is indicative of underlying periodic dynamics in the time series. In recurrence quantification analysis, the distribution of these features is quantified to derive metrics descriptive of underlying processes in the elemental time series.

**Table S1. Features derived from recurrence and/or cross-recurrence quantification analysis of elemental time series data.**

Feature	Algorithm	Interpretation
Recurrence Rate (RR)	$RR = \frac{1}{N-k} \sum_{j=1}^{N-k} R(i, j)$ <p>Where <math>k</math> is the distance of a given diagonal line from the central diagonal in a square recurrence or cross recurrence matrix, <math>N</math> is the total number of diagonal processes, and <math>R(i, j)</math> is the probability of recurrence at a given coordinate <math>i, j</math>.</p>	This feature measures the ratio of recurrent points, typically indicated in black, vs. non-recurrent points (white space) in a recurrence matrix or cross-recurrence matrix. This measure reflects the incidence of repetition in each signal. In these analyses, threshold functions were set adaptively to yield an $RR=0.1$ in all analyses; as $RR$ was therefore a constant, this feature was excluded from analysis.
Determinism (DET)	$DET = \frac{\sum_{l=l_{min}} lP(l)}{\sum R}$ <p>Where <math>l</math> is the length of diagonal lines in the recurrence matrix, <math>P(l)</math> is the number of lines with length equal to <math>l</math>, and <math>R</math> is the number of recurrence points in the full recurrence plot.</p>	This feature measures the ratio of diagonal points to non-diagonal points in a recurrence matrix / cross-recurrence matrix. This reflects the prevalence of periodic processes in the signal.
Mean Diagonal Length (MDL)	$MDL = \frac{\sum_{l=l_{min}}^N lP(l)}{\sum_{l=l_{min}}^N P(l)}$ <p>Where <math>N</math> is the number of diagonal lines, <math>l_{min}</math> is the minimum specified length of diagonal lines (in this application, 2), <math>l</math> is the length of a given diagonal line, and <math>P(l)</math> is the histogram of all diagonal lines.</p>	This feature measures the average length of diagonal lines in a recurrence matrix / cross-recurrence matrix. This measure reflects the duration of periodic processes in the signal.
Maximum Diagonal Length (LMAX)	$LMAX = \max(P(l))$ <p>Where <math>P(l)</math> is the histogram of diagonal lines in a recurrence or cross recurrence matrix.</p>	This feature measures the longest diagonal line in a recurrence matrix / cross-recurrence matrix. This measure reflects the duration of the most persistent periodic process.
Shannon Entropy in Diagonal Length (ENT)	$ENT = - \sum_{l=l_{min}}^N p(l) \ln p(l)$ <p>Where <math>l</math> indicates the length of a diagonal line, <math>l_{min}</math> indicates the minimum length of diagonal lines (in this application, 2), <math>N</math> is the total</p>	In a recurrence matrix / cross-recurrence matrix, this feature measures the predictability of the distribution of diagonal line length. This can be interpreted as a measure of complexity involved in periodic processes in the signal.

	number of diagonal lines, and $p(l)$ is the probability distribution of diagonal lines.	
Trend in Recurrences (TREND)	$TREND = \frac{\sum_{i=1}^N (i - \frac{N}{2})(RR_i - \langle RR_i \rangle)}{\sum_{i=1}^N (i - \frac{N}{2})^2}$ <p>Where <math>N</math> is the number of diagonal lines parallel to the maximal diagonal in a square recurrence or cross-recurrence matrix, <math>i</math> is a diagonal line, <math>RR_i</math> is the diagonal-wise recurrence rate, and <math>\langle RR_i \rangle</math> is the average diagonal recurrence rate.</p>	This feature measures the relative density or sparseness of a recurrence matrix at the matrix edges relative to the center. This measure reflects the trend in the signal to depart from stationarity.
Laminarity (LAM)	$LAM = \frac{\sum_{v=l_{min}} v P(v)}{\sum R}$ <p>Where <math>v</math> is the length of vertical lines the recurrence or cross-recurrence matrix, <math>l_{min}</math> is the minimum length of vertical lines (in these applications, set to 2), and <math>P(v)</math> is the histogram of vertical line lengths. <math>R</math> refers to the sum of all recurrence points; as such, this function essentially constructs a ratio of recurrence points which contribute to vertical lines relative to all recurrence points.</p>	This feature measures the ratio of vertical lines in a recurrence or cross recurrence matrix relative to points that do not comprise vertical lines. This measure characterizes the prevalence of periods of stability in a given signal.
Trapping Time (TT)	$TT = \frac{\sum_{v=l_{min}} v P(l)}{L}$ <p>Where <math>v</math> is the length of vertical lines the recurrence or cross-recurrence matrix, <math>l_{min}</math> is the minimum length of vertical lines (in these applications, set to 2), and <math>P(l)</math> is the histogram of vertical line lengths. <math>L</math> in this case refers to the frequency of all vertical lines meeting <math>l_{min}</math> criteria; thus, in essence this measure reflects the mean vertical line length.</p>	This feature measures the distribution of temporal intervals between vertical lines in a recurrence matrix or cross recurrence matrix. This measure provides the mean interval observed between periods of signal stability.
Maximum Vertical Line Length (VMAX)	$VMAX = \max (P(v))$ <p>Where <math>P(v)</math> indicates the distribution of vertical line lengths.</p>	This feature measures the maximum length in the distribution of vertical line lengths observed in a recurrence matrix or cross-recurrence matrix. This measure reflects the peak duration of stable processes in a given signal.
Shannon Entropy in Vertical Line Lengths (VENT)	$VENT = - \sum_{v=l_{min}}^N p(v) \ln p(v)$ <p>Where <math>l</math> indicates the length of a vertical line, <math>l_{min}</math> indicates the minimum length of vertical lines (in this application, 2), <math>N</math> is the total number of vertical lines, and <math>p(v)</math> is the probability distribution of vertical lines.</p>	This feature measures the predictability of the distribution of vertical line lengths in a recurrence or cross-recurrence matrix. This measure reflects the complexity in the distribution of intervals of signal stability.

Mean Recurrence Time (TMRP)	$TMRP = \frac{\sum_{w=1}^N wP(w)}{\sum_{w=1}^N P(w)}$ <p>Where <math>w</math> is the length of vertical white lines, <math>N</math> is the total number of vertical white lines, and <math>P(w)</math> is the frequency distribution of vertical white line lengths.</p>	This feature measures the mean duration of intervals between vertical lines indicative of signal stability in a recurrence or cross-recurrence matrix.
Shannon Entropy in Recurrence Time (WENT)	$WENT = - \sum_{w=1}^N p(w) \ln P(w)$ <p>Where <math>w</math> is the length of vertical white lines, and <math>P(w)</math> is the frequency distribution of vertical white lines.</p>	This feature measures the predictability of the distribution of intervals separating vertical lines indicative of signal stability in a recurrence or cross-recurrence matrix.
Number of the most probable recurrence time (NMPRT)	$NMPRT = \max (P(w))$ <p>Where <math>P(w)</math> indicates the histogram of the vertical white line distribution</p>	This feature measures the maximum value of the histogram of recurrence times. This measure indicates the most probable interval between periods of signal stability.

### III. Model Performance – Confusion Matrix

**Table S2. Confusion matrix: frequencies of predicted diagnoses and actual diagnoses.**

		Reference Test - DSM-V		
		Positive	Negative	Total
Model Predictions	Positive	27	17	44
	Negative	1	52	53
	Total	28	69	97

Table cells reflect the number of predicted diagnostic outcomes relative to actual diagnostic status as determined by the DSM-V in the holdout validation dataset.

#### IV. Participant Comorbidity Characteristics

**Table S3. Comorbid diagnoses in participants with and without positive diagnosis for ASD.**

Disorder	ASD	Control
<i>ADHD</i>	26	21
<i>Intellectual</i>	16	4
<i>Developmental</i> <sup>1</sup>	30	17
<i>Psychiatric</i> <sup>2</sup>	16	6
<i>Genetic</i> <sup>3</sup>	9	5
<i>No comorbidity</i>	123	279

<sup>1</sup>Developmental Delay, Learning Disorder, Language Disorder

<sup>2</sup>Anxiety, Depression, OCD, Phobia, PTSD, Tics, Tourette's

<sup>3</sup>Genetic Syndrome, PMS, DDX3X, FOXP1, ADNP

Cells reflect counts (n) of participants with positive diagnoses in each broad diagnostic category; note diagnoses are non-exclusive, such that a patient might receive multiple diagnoses.