bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Sensing Echoes: Temporal misalignment as the Earliest Marker of Neurodevelopmental Derail

Elizabeth B Torres[1,2,3,*], Thodoris Bermperidis[1], Hannah Varkey[1], Joe Vero[1], Eric London[4], Ha Phan[4], Phyllis Kittler[4], Anne Gordon[4], Simon Schafer[5], Fred Gage[5], Terrence Sejnowski[5]

[1] Rutgers University Psychology Department

[2] Rutgers University Center for Cognitive Science

[3] Computational Biomedicine Imaging and Modelling, Computer Science Department

[4] New York State Institute for Basic Research in Developmental Disabilities

[5] The Salk Institute for Biological Studies

Abstract

Sensory transduction and transmission delays operate and propagate along different time scales. From microseconds in the auditory domain, to hundreds of milliseconds in the visual, and kinesthetic domains, the brain must successfully align disparate delays arising from endogenously self-generated streams of motor and visceral sensorial information, with exogenous sensory inputs. To produce a cohesive response to environmental goals, constantly explore, adapt, and develop a sense of simultaneity, the brain must resolve this major feat and compensate for excessive delays in any sensory modality. Disruption in these processes may lead to altered perception of the self and others, and inadvertently affect social interactions. But how early such issues may emerge and be reliably detectable, remains a challenge. Here we assess in neonates, the transmission latencies of a sound wave that travels from the cochlear nerve to the brainstem on its way to the primary auditory cortex. Already at birth, we find systematic and cumulative delays in the propagation of this wave in neonates that later received a diagnosis of autism. Furthermore, we discover that the distributions of such temporal delays have far narrower bandwidth than those from neonates who did not receive the autism diagnosis. We identify associated codependent genes' networks and define a reliable marker of neurodevelopment derail, detectable at birth. Under the precision autism model, we propose that the brainstem contains an endogenous clock anchoring and aligning disparate timescales critical for the emergence and maintenance of congruent percepts of the self and others.

For Peer Review

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To control our bodies in motion, our brain must rely on internal representations of space and time [1-5], that from early stages of life, endogenously self-emerge and are updated, from the intrinsic perspective of the organism [4-6]. In due time, the human infant will acquire the ability to shift this perspective and think of himself (and others) from the imaginary vantage point of a disembodied observer, who learned to theorize about and timely adapt to the mental and physical dynamics of others in a social scene [6]. Yet to get there, basic foundational elements involving dynamic processes unfolding at different time scales ought to be in place (Figure 1A). The person must have developed spatio-temporal maps of the self (sensory motor percepts.) Such maps ought to be congruent with those of others, to effectively establish communication systems and from an early age, accurately and timely conceptualize, distinguish, and predict real from imaginary self.

Building blocks of social interactions begin developing since conception and mature throughout early infancy [7-9]. Indeed, as altricial mammals, the somatic sensory motor structures of human neonates require further maturation after birth [10-13]. In contrast, the brainstem's structures will already (at birth) have reached a level of maturation to enable communication across still-developing central, peripheral, and autonomic neuronal networks [14-16]. The success of such processes facilitates back-and-forth communication along efferent and afferent nerve fibers, connecting the central, the peripheral, and the autonomic nervous systems, to help compensate for transduction and transmission delays in somatic, sensory, and motor processes [9, 10, 12, 13] (Figure 1B). Every aspect of behavior depends on the success of these maturation processes resulting in the emergence of properly aligned and predictable temporal delays. These are ultimately destined to endow the person with a sense of simultaneity and a cohesive percept of the world, to succeed at purposeful actions and social, affective interactions with others [6].

In the nascent nervous system, disparate latencies of information traveling across the different networks that cross through the brainstem, emerge from delays incurred at different stages of sensory transduction and transmission processes. These are quantifiable using standard *ms* timescales at the sites of the brainstem where the propagating signal arrives (Figure 1C). The intactness of signal's transduction and transmission depends on interactions and co-dependencies among a multiplicity of genes driving the formation, maturation and maintenance of brainstem structures and interconnected cortical-fronto-brainstem-cerebellar neural networks (Figure 1D).

Multiple levels of signal processing contribute to variability in these delays. From molecules, enabling gap junctions and the timely opening and closing-of channels at the synapse, to pre- and post-synaptic communication at the dendrites, to axonal communication across neurons and neuronal networks at the systems' level, each layer adds variations to these latencies. In terrestrial mammals these delays are also contributed by latencies across the muscle junctions and joints, as the body moves with many degrees of freedom (DOF) and incurs in inertial delays at the periphery [10, 12, 13]. As some DOF are devoted to sensory goals and error correction,

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

others are devoted to spontaneous exploration, thus distinguishing fundamentally different types of delays to learn, manage, and predict [17-19]. Fluctuations in such delays are also stochastic streams of reafferent feedback conducive of probabilistic codes that when appropriately maturing, can anticipate the sensory consequences of impending actions and thus compensate for the delays [6].

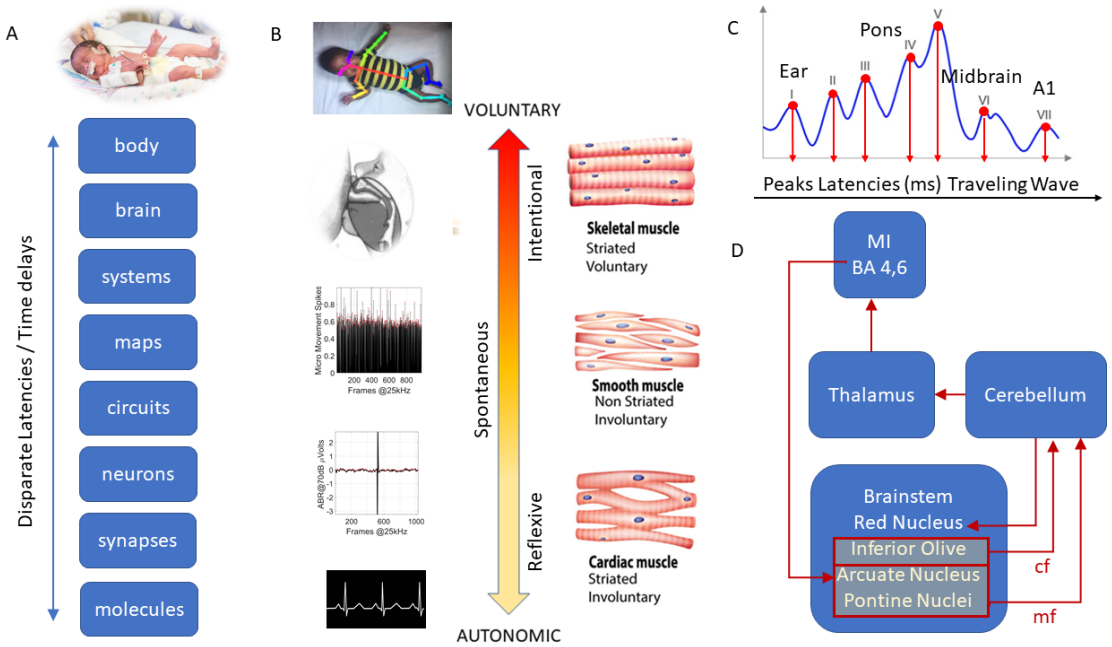


Figure 1. The altricial human mammals whereby brainstem development precedes maturation of somatic-sensory-motor control. (A) Disparate latencies in signals' transduction and transmission incur in variable time delays that the brain needs to align, to form coherent percepts of the self and of others in the world, to build appropriate frames of reference to perform coordinate transformations from disparate exogeneous and endogenous sensory signals and to compensate for such variable delays and be able to navigate these sensory motor spaces mentally and physically. Different orders of magnitude contribute to these variable delays and bring stochasticity to each level, from molecules to complex behaviors of bodies in motion. (B) The Precision Autism platform integrating individual information from different layers of the knowledge base, from clinical self-reports to precision behavioral analyses to omics [17]. At the behavioral and genomics level we use micro-movements to characterize fluctuations in signals (gross data otherwise discarded) harnessed with non-invasive, off-the-shelf sensors and using open access big data connecting genes networks and fundamental tissues supporting the nervous system. Phylogenetically ordered taxonomy of neuromotor control is supported by fundamental muscle types linked to basic gene pools for survival. (C) ABR latencies have been characterized in neonates and during childhood. The timings (*ms*) of peaks I-VII are registered as the signal propagates through the brainstem, to test the intactness of cranial nerve VIII starting at the ear, following through the pons, the midbrain and finally arriving at the primary auditory cortex. This is done at clinical settings to estimate possible developmental delays for different populations. (D) Known neuroanatomical structures of the cortical-fronto-cerebellar-brainstem loops have been identified as disrupted in autism (cf, climbing fibers, mf, mossy fibers, schematics adapted from [20].)

The brainstem is bound to partake on such coding, as it is a conduit gating information flow from and to centers that control fine and gross motions from head to toe, throughout awake and sleep cycles [20]. These include as well structures connecting autonomic self-regulation and emotions [16]. Furthermore, the development of the brainstem monoaminergic circuitry is

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

fundamental to later cortical development [21]. It is also essential for the tonotopic organization of the auditory system [22, 23], enabling it to receive sensory-evoked signals and produce orderly measurable responses [24], Figure 1C. As such, early disruptions in the brainstem has been hypothesized to play a causal role in neurodevelopmental derail, inevitably leading to the umbrella-type autism spectrum disorders diagnosis [25]. Yet, how disturbances at different levels may lead to differences in latencies as signals propagate through different brainstem regions, and what types of genomic interactions may be relevant to such disturbances, remain unknown.

At embryonic stages, the auditory sophistication of the fetus and later of the newborn, has been noted [15, 26, 27]. It has been reported that the human auditory system undergoes central maturation before birth, with synchronous axonal myelination across the proximal cochlear nerve and brainstem, between the 26th and 29th fetal weeks of gestation [11]. During this prenatal period, these detectable myelin sheaths have been reported along the brainstem structures including axons of the cochlear nerve, trapezoid body, lateral lemniscus, dorsal commissure of the lemniscus, commissure of the inferior colliculus and brachium of inferior colliculus. After the 29th fetal week of gestation, subsequent density of myelination are thought to be a factor in the steady decrease of auditory brainstem evoked response (ABR) latencies reported in neonates [11].

As the brainstem provides such an important foundation for proper neurodevelopment, it seems critical to study stochastic interactions and co-dependencies in genomic pools contributing to these early, highly dynamic maturation processes. These include processes mediating oligodendrogliosis, transitioning to incipient myelin sheaths, and then to definite myelination ensuring proper signal transmission latencies, along the continuous streams of sensory transduction [11]. Postnatal axonal myelination has been reported to radically increase the rate of synchronicity of transmission throughout the auditory pathways, including the cochlear nuclei, superior olive, lateral lemniscus, and inferior colliculus [28].

Here we posit that such early processes create fundamental building blocks of social interactions mediated by the nervous systems. They help anchor information from disparate time scales onto a unifying endogenous clock. What happens then, if this clock is mis calibrated and the typical decrease in response latencies along the brainstem is missing at birth and/or significant delays persist throughout early infancy? Can we identify genomic networks possibly involved in these early neurodevelopmental processes?

Results

We assessed the ABR waveforms of 54 neonates broken down in Table 1 according to different categories (Full Term FT, Neonatal Intensive Care Unit NICU, Pre-Term PT and Well Nursery Babies WNB.) Within those, we had a group who went on to receive a diagnosis of autism spectrum disorder (asd) and another group that did not, (non asd.) We had multiple trials (L1, L2, L3) each one comprising different sound levels 70, 75, 80 dB. Total trials are 108 including,

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

66 with some repeated trials, (full set of 3 levels each), 54 L1; 8 L2 and 4 L3. We had 54 babies for personalized analyses with a unique trial L1 encompassing all 3 dB levels and pooled across all trials for population analyses addressing distinctions in ABR amplitudes and latencies for each baby type.

ABR TEST 1 (L1)				
	NICU PT # babies	**NICU FT** # babies	**WNB (FT)** # babies	**TOTAL** # babies
ASD	16	2	6	24
Non ASD	16	3	11	30
TOTAL	32	5	17	54
ABR TEST 2 (L2)				
ASD	3	1		4
Non ASD	3	1		4
TOTAL	6	2		8
ABR TEST 3 (L3)				
ASD	4			4
Non ASD				
TOTAL	4			4

Table 1. Number of babies included in the study and sets L1, L2, L3 of repeated trials. Each trial contained three levels of sound spanning 70, 75, 80dB from which the full waveform was saved.

We also had access to additional trials from 14 FT babies (2 FT NICU babies) and 34 NICU PT babies. However, these data sets were incomplete (*e.g.*, missing data from a dB level or two) or too noisy. As such, these data were not included in the present analyses.

In addition to the full waveform data (μV), we also had access to the latency data (ms) from 47 trials including the following neonates (13 FT, 11 FT NICU, 2 FT nonNICU with 8 asd and 34 PT with 30 asd). As with the full waveform, we had multiple trials in each set of responses to three dB levels. Pooling across all neonates and trials, and using bootstrapping techniques, we built a larger dataset (above 100 measurements) to interrogate fluctuations in latencies and various features of the full waveforms. Furthermore, we used a personalized approach to also examine the individuals in the cohort, in search of self-emerging patterns. We here compare the ABR latencies at ms time scale along with the fluctuations in the waveform's peaks' amplitude (μV) and peaks' width (ms) reflecting these responses. Although such waveform's features are seldom examined, we introduce new methods to assess several empirical parameterizations related to the micro-fluctuations of the waveforms' peaks features (amplitude, prominence, and width). These micro-fluctuations are commonly discarded as gross data or smoothed out as noise. Yet, we find that they automatically separate the asd from the non-asd groups and clearly differentiate across dB levels evoking the responses and differentiating between pre-term and full-term neonates.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

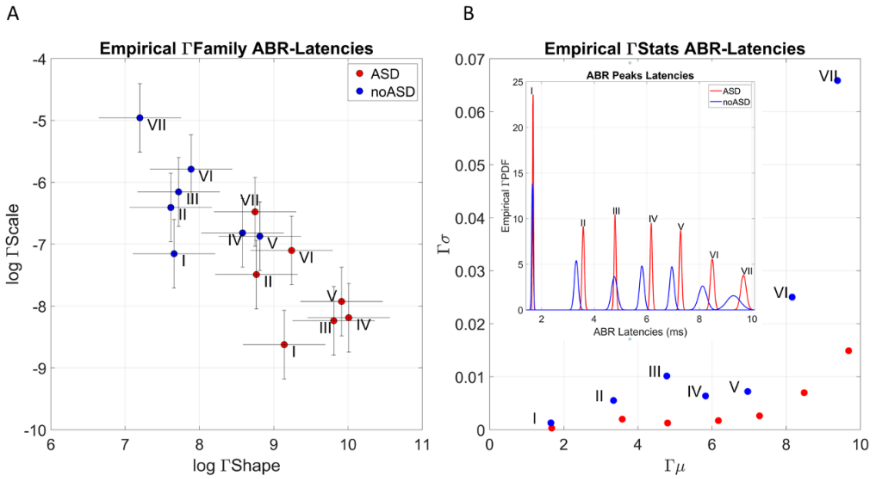


Figure 2 Different families of empirically estimated probability distribution functions (PDFs) of the evoked response potential latencies characterize the asd *vs*. no-asd neonates across the I-VII regions. (A) Log-log Gamma parameter plane spanned by the Gamma shape and scale parameters show for each brainstem region a complete separation between cohorts. Points represent the PDF's shape and scale (dispersion), with 95% confidence intervals. The asd-neonates group in the region of distributions with lower dispersion and more symmetric shapes. (B) The empirically estimated Gamma moments (mean and variance) emphasize the differences in (A). The no-asd neonates have broader range of latency variability and faster timings than the asd-neonates. Inset shows empirical PDFs superimposed for both cohorts, highlighting the shifts in timings (*ms*) as the evoked response signals propagate along the seven regions.

In a rare opportunity, we had access to data from three distinct click levels at 70, 75 and 80 dB (see methods), to systematically track the propagation times of the response waveform and assess the intactness of the cranial nerve VIII. We used seven standard regions of interest, a subset of them previously characterized in neonates [29]. They are depicted in Figure 1B. Yet, we did not average these values across the pool of neonates under consideration, using an *a priori* assumed theoretical distribution (as in *e.g.*, [29]). Instead, we empirically estimated the best continuous family of PDFs fitting these data. We found distributions of latencies with heavy tails. These were well fit by the continuous Gamma family of probability distributions using maximum likelihood estimation (MLE). As such, empirically obtaining the Gamma shape and scale parameters and plotting them on the Gamma parameters' plane with 95% confidence intervals, provided an overall view of the stochastic signatures of these two groups of neonates as latencies propagated across brainstem regions I-VII. Furthermore, the Gamma PDFs thus obtained for each set of latency values revealed fundamental shifts in stochastic signatures between the babies who received the autism diagnosis and those who did not.

Figure 2A shows the 14 points representing the empirically estimated shape and scale Gamma parameters on a *log-log* plane for the seven sites (I-VII) corresponding to the non-asd neonates, and those corresponding to the asd-neonates. We note that under equal number of measurements (from bootstrapping), the asd babies have far lower dispersion in their distributions, indicating a much narrower bandwidth of latencies than the non-asd babies. Furthermore, the densities of the distributions of asd babies are shifted to more symmetric

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

shapes, unlike the non-asd babies having more skewed distributions. These distributions with heavy tails emerge from broader ranges of ABR latencies at each of the I-VII stages. Pairwise, comparing at each stage the non-asd with the asd stochastic signatures, there is no overlap between these probability distributions representing the two groups. Using the two-sample Kolmogorov-Smirnov test, we found significant *p-values* ($p<0.01$) pairwise between the two groups. Supplementary figure 1 A-G shows the frequency histograms from the raw data whereby, using bootstrapping techniques to draw from the larger set equal number of points to the smaller set, we estimate distributions with equal number of participants, to compare between cohorts and build a distribution of *p-values* derived from the comparisons. Supplementary Figure 1H shows the pairwise statistical comparison thus obtained with two asterisks representing $p<0.01$ and one asterisk representing $p<0.05$. Furthermore, using the earth mover's distance (EMD), we quantify, pairwise, the similarity across the 14 frequency histograms (7 regions from asd *vs*. 7 regions from non-asd neonates.) Supplementary Figure 2 shows the matrix entries color-coded, as indicated by the colorbar depicting the normalized EMD values. These data were used to produce empirical estimates of the PDFs depicted in Figure 2B inset and offering a view of the systematic shifts (delays) in the asd group. Figure 3 plots each of the PDFs at each of the I-VII stages, suggesting cumulative delays (in *ms*). The figure visualizes the systematic narrower ranges in latencies for the asd group (despite having the highest number of babies.)

The cumulative effect across these sites of the brainstem revealed 1.74*ms* difference, with 0.40*ms* as the longest local delay in the region VII, while region III had the shortest delay 0.05*ms*. Table 1 summarizes all latencies (*ms*) across the seven points of interest for each participant type. The asd neonates show a net cumulative delay of 8*ms vs.* 7.62*ms* in the non-asd neonates.

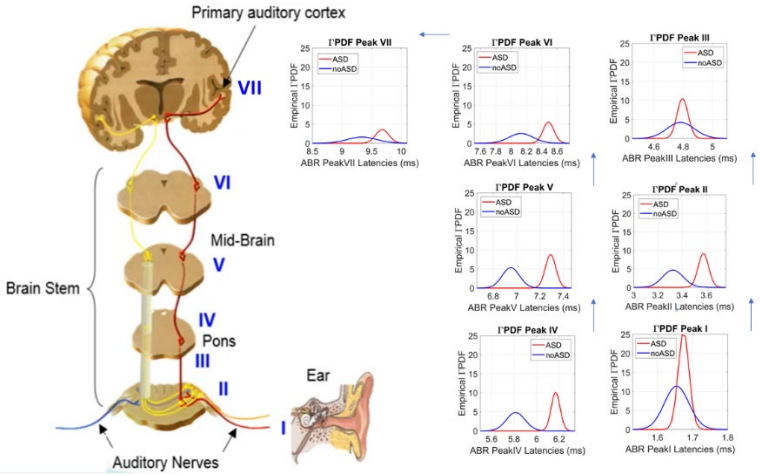


Figure 3 Empirically estimated family of continuous Gamma probability distributions characterizing the asd and non-asd cohorts of neonates. Each plot compares the two PDFs for a site (I-VII) and clearly demonstrates the (cumulative) delays in the non-asd group along with the narrower bandwidth of values. Brainstem schematics is included for visualization purposes.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Regions	I	II	III	IV	V	VI	VII	Cum Σ
asd	1.68	3.58	4.80	6.17	7.29	8.48	9.68	
Δ asd		1.90	1.22	1.37	1.12	1.19	1.20	8.00
non asd	1.66	3.33	4.78	5.83	6.95	8.11	9.28	
Δ non asd		1.67	1.45	1.05	1.12	1.16	1.17	7.62
Interpeak Δ (non asd-asd)	0.02	0.25	0.02	0.34	0.34	0.37	0.40	1.74

Table 1 Peak latencies for each site in the two cohorts and inter-peak latency differences. Last column shows the cumulative sum of the delays and the appreciable asd neonates' delay of 1.74 *ms* (considering that sound processing is generally agreed to be on a μs time scale)

This information is also shown in Supplementary Figure 3, according to the individual body weight, BW and sex for each group. There we see that non-asd females tend to have higher latencies in VI and VII sites, and overall higher cumulative latencies than non-asd males. However, there were insufficient female neonates in the asd group to have any meaningful trend with statistical power. As such, these results are skewed toward asd males.

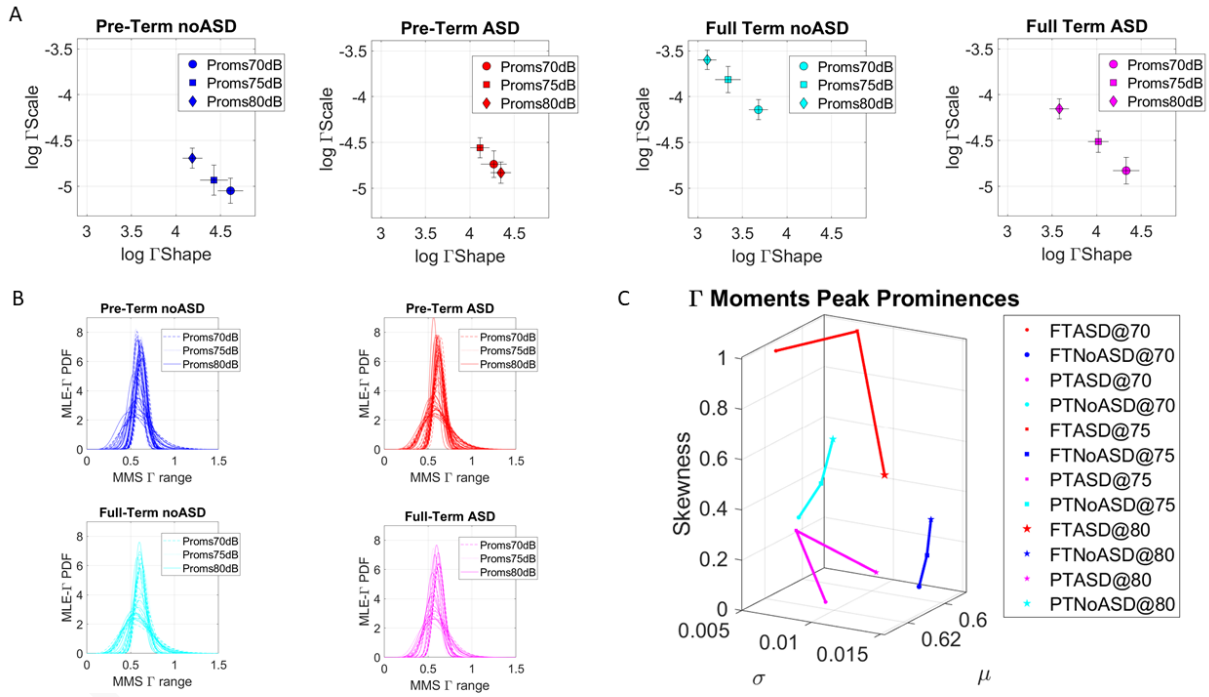


Figure 4 Fluctuations in peaks' prominences separate dB levels in asd and non-asd pre-term and full-term neonates, and automatically distinguish for each level, the different groups. (A) Log-log Gamma parameter plane contains the stochastic signatures for each dB level (with 95% confidence intervals). These levels are different for each of the asd *vs*. non-asd cases and pre-term *vs*. full-term cases, with the later showing maximal separability between asd and non-asd neonates. (B) Corresponding empirical Gamma PDFs obtained for each baby that went into the computation of the Gamma shape and scale group parameter. (C) Empirical Gamma moments spanning parameter space along the x-axis (mean), y-axis (variance), z-axis (skewness) and using the kurtosis proportional to the marker's size. The stochastic signature of each dB level connects to the next level via a line representing a path in probability space. This path clearly marks differentiation in the transitions of the signal from one dB level to the next in asd *vs*. non-asd neonates.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Waveform's prominences separate asd from non-asd in full-term vs. pre-term babies

The waveforms reflecting the brainstem responses have not been analyzed in their full extent, as only the peaks' averages had been of interest to prior research [29, 30]. Here, instead of averaging the peaks and smoothing out as noise the fluctuations across the dataset, we obtained the full waveform inclusive of activity prior and after the peak. We examine the variations in the fluctuations of their widths, their prominences, and their full amplitudes (defined in the Methods Figure 11). Pooling across the autism and non-autism cohorts, we found detectable differences in the empirically estimated Gamma PDFs for each of the features.

Most notably, we identified the prominences as the feature revealing the largest differentiation in PDFs between these two cohorts, maximally separating asd from non-asd full-term neonates across the dB levels. This waveform feature also systematically differentiated asd pre-terms from non-asd pre-terms (Figure 4A shows the separation on the log-log Gamma plane, while Figure 4B does so for the empirical PDFs. Figure 4C shows the corresponding empirical Gamma moments along a path connecting the signatures for each dB level's evoked response. We can automatically separate these cohorts along all three dB levels under consideration and distinguish pre-term from full-term neonates according to the asd *vs*. non-asd subtype. Supplementary Figures 4-6 show results for the peaks' amplitude, proms, and widths, while Supplementary Figure 7A depicts the EMD matrices with entries denoting pairwise similarity in amplitude and inter-peak-interval latencies of the full waveform, while 7B shows the results from non-parametric pairwise comparisons of the differences in peak amplitude values for each subgroup of the 70-75-80dB -prominences (left) and amplitudes (right). Here the differentiation of three dB levels in full-term *vs.* pre-term asd neonates are statistically indistinguishable. This means that the fluctuations in the differences in peaks' prominences across the levels cannot separate between pre-term and full-term asd babies. All other pairwise non-parametric comparisons (Wilcoxon rank sum test) yielded statistically significant differences for these three dB levels' differentiation at the *0.05* alpha level.

A study of the stochastic ranges between the minimum and maximum shape and scale values is depicted in Supplementary Figure 8. This analysis also confirmed the differences between groups, along with self-emerging clusters separating asd *vs.* non-asd individual neonates in Supplementary Figure 9. Supplementary Figure 8 represents a parameter space spanned by the skewness of the distribution of peaks' width along the x-axis, the variance of the distribution of peak's amplitude along the y-axis and included the body weight (BW) at the first visit (panel A), or the estimated gestational age (EGA) at birth (panel B) along the z-axis. Each point in this parameter space represents one baby (personalized stochastic signatures) along each estimated family of probability distributions spanned by the waveforms' features. The graph shows automatically emerging clusters of babies. This clustering reveals that low BW and low estimated gestation age, EGA are not predictors of autism. The cohort of asd neonates had babies across low and high values of both parameters. Supplementary Figures 10-12 further studies the relations of EGA, BW, NSR and empirical distribution skewness. Indeed, Methods

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 10 reveals the non-linear relations between BW and EGA along with the full demographics of the study.

Identification of Genomic Drives Underlying Disparate Latencies – Marginal Distributions

Given that the asd cohort was composed by neonates who received the asd diagnosis, and that there is a strong genetic component to asd, we downloaded the set of 906 autism-associated genes from the open access databank maintained by the Simons Foundation Autism Research Initiative (SFARI). As in our recent work [31], we also performed a search in the disease associated genes network (DisGeNET) inclusive of 110 genes linked to multiple brainstem disorders. These disorders ranged across auditory neuropathy spectrum disorder, auditory neuropathy, abnormal auditory evoked potential, central auditory processing disorder, auditory perceptual disorders, absent brainstem auditory responses, auditory inattention, auditory neuropathy autosomal dominant 1, auditory neuropathy nonsyndromic recessive, prolonged brainstem auditory evoked potentials, and hyperacusis. We then examined these genes' expression in 54 human tissues, using the Genotype-Tissue Expression (GTEx) consortium atlas of genetic regulatory effects across human tissues. These open access data involve human RNA-seq, expressed in Transcripts Count Per Million (TCM) [32] (see methods for further details and [31].

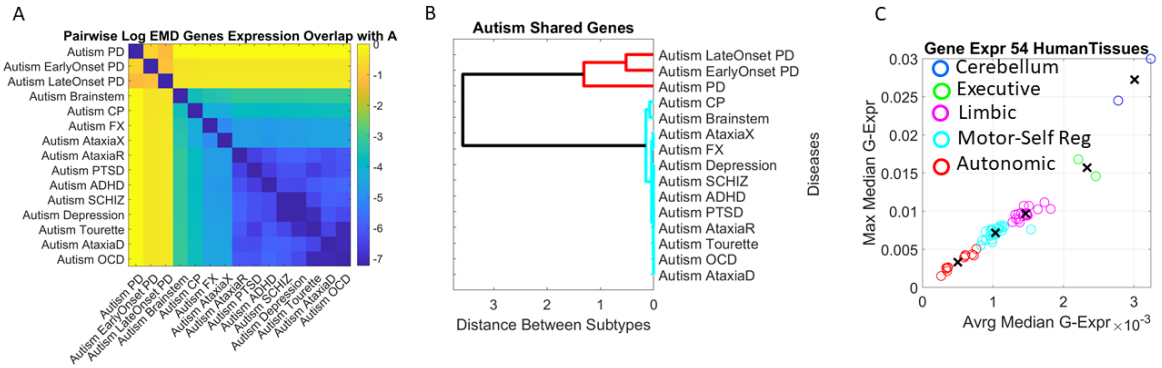


Figure 5 Localizing the genes associated to brainstem disorders shared with the autism associated genes in SFARI on the space of genes from neurological and neuropsychiatric disorders, also intersected with the SFARI genes. (A) The genes' expression of 21,033 genes across 54 human tissues in GTEx. These were obtained from genes shared between the autism associated SFARI set and each of 15 sets of genes associated to nervous systems disorders (inclusive of brainstem disorders). The normalized gene expressions across the 54 tissues were compared pairwise using the EMD on the frequency histograms (marginal distributions) of each set. The log of the normalized EMD value is then used to rank the distances in a colormap. (B) A dendrogram graph showing the automatic grouping of the intersecting sets related to Parkinson's Disease PD (general, early onset and late onset) and other neurological and neuropsychiatric disorders (Cerebral Palsy-CP, Fragile X-FX, Ataxias dominant and recessive, Post Traumatic Stress Syndrome-PTSD, Attention Deficit Hyperactivity Disorder-ADHD, schizophrenia, depression, Tourette's and Obsessive-Compulsive Disorder-OCD). Notice that the set of genes shared between brainstem disorders and SFARI autism is closest to the set of genes shared between CP and SFARI autism. Furthermore, the genes shared between SFARI autism and depression, SFARI autism and schizophrenia, SFARI autism and ADHD cluster together. (C) Automatic clustering using K-means of the averaged normalized median gene expression *vs.* the maximal median gene-expression on the 54 tissues involving the pool of genes shared between SFARI autism and all disorders under consideration reveal the cerebellar hemisphere and cerebellar cortex as a unique group. Other clusters included tissues with the functions described in the legend (see main text for more details.)

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We obtained 72 genes common to both the above-mentioned brainstem disorders and the SFARI autism associated genes. Furthermore, we intersected the genes associated with each of the other 14 neuropsychiatric and neurological disorders and the SFARI autism genes. Then, we ranked them by similarity distance (EMD metric) based on the marginal distributions of gene expression across the 54 human tissues of GTex (see methods Figure 13). Figure 5A shows the colormap built from the pairwise (normalized) probabilistic distances using EMD. Figure 5B shows the dendrogram graph depicting a binary tree with emerging clusters based on genes' expression on the 54 human tissues (cophenetic correlation coefficient 0.98). This graph localizes the genes common to brainstem disorders and SFARI autism closest to those common to cerebral palsy (CP) and SFARI autism. Figure 5C shows the clusters of tissues upon K-means using 5 groupings of points representing the average *vs.* maximal median expression per tissue. These tissues self-grouped according to various functions labeled in the legend. Specifically, top cluster (blue) with highest expression included the cerebellum and cerebellar hemisphere important for motor coordination, adaptation, learning and overall, for timing, internal models of motor dynamics and cognitive aspects of motor control [33]. The second cluster (green) based on genes' expression includes the frontal cortex and the pituitary gland important for executive function and regulatory control, respectively. Following this cluster in gene expression (magenta), we see the brain anterior cingulate cortex, the caudate and nucleus accumbens of the basal ganglia, the brain cortex, the hypothalamus, and the spinal cord cervical C1. These tissues are important for overall motor control and self-regulation. The next cluster (cyan) involves the limbic system, with the amygdala and hippocampus, but also the putamen of the basal ganglia and the substantia nigra. These tissues are involved in emotion, memory, navigation, and motor functions. The last cluster (red) involves autonomic functions via the heart atrial appendage and left ventricle as well as overall function of the muscle skeletal system.

Identification of Genomic Drives– Dynamic Codependent Network Interactions

These analyses gave us a sense of where brainstem-related issues phenotypically stand in probability space, involving frequency histograms (marginal distributions of gene expression.) However, we further wished to understand the dynamic developmental role of the subset of genes shared between the SFARI autism set and the brainstem disorders. To that end, we examined neighboring genes' interactions combining probabilistic graphical models and maximal spanning tree. We identify 72 genes with variable degree of conditional dependencies. We then track their dynamically changing expression paths during early embryonic stages, using a published data set of human embryonic stem cells (hESCs) evolving to become neurons [34].

Using undirected graphical models whereby any two nodes without a connecting edge are conditionally independent (statistically independent, given all other nodes), we tracked those nodes with an edge, *i.e.*, conditionally dependent. In these dependency graphs, we tracked their dynamically changing co-dependencies (see methods) measured through mutual information (MI). We estimated MI from their gene's expression across hundreds of cells, on days 12, 19, 40

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and 54, as the hESCs evolved into neurons. We then used Chow-Liu (maximal spanning) trees [35] to study connectivity patterns (see analytical pipeline in methods Figure 14) as they shifted in time and formed hubs of varying degrees.

The factorization of the joint probabilistic behavior of the network determined the significance of each gene in the network as the node's degree. The genes with highest degree are thus the hubs with maximal interdependencies, driving the evolution of its neighboring genes in the ESCs, as they develop into neurons. As our graph is undirected, we do not know exactly what role(s) these *hub* genes with high statistical interdependencies may play (*e.g.*, they may be essential for the functioning of their neighbors, or their neighbors' co-expression may drive that gene, *etc*.) Regardless of the causal role(s) between node and neighbors, we can certainly track these hubs dynamically and qualitatively describe, using the literature, their possible roles in early embryonic neural development. This type of description will help us understand their possible roles in early neurodevelopmental derail. On the other hand, genes with low degrees have fewer co-dependencies, which could potentially mean that their role is separate from the role of clusters of highly interconnected genes. Perhaps their function is independent and not part of a large biochemical cascade leading from embryonic pluripotent states to neuronal states.

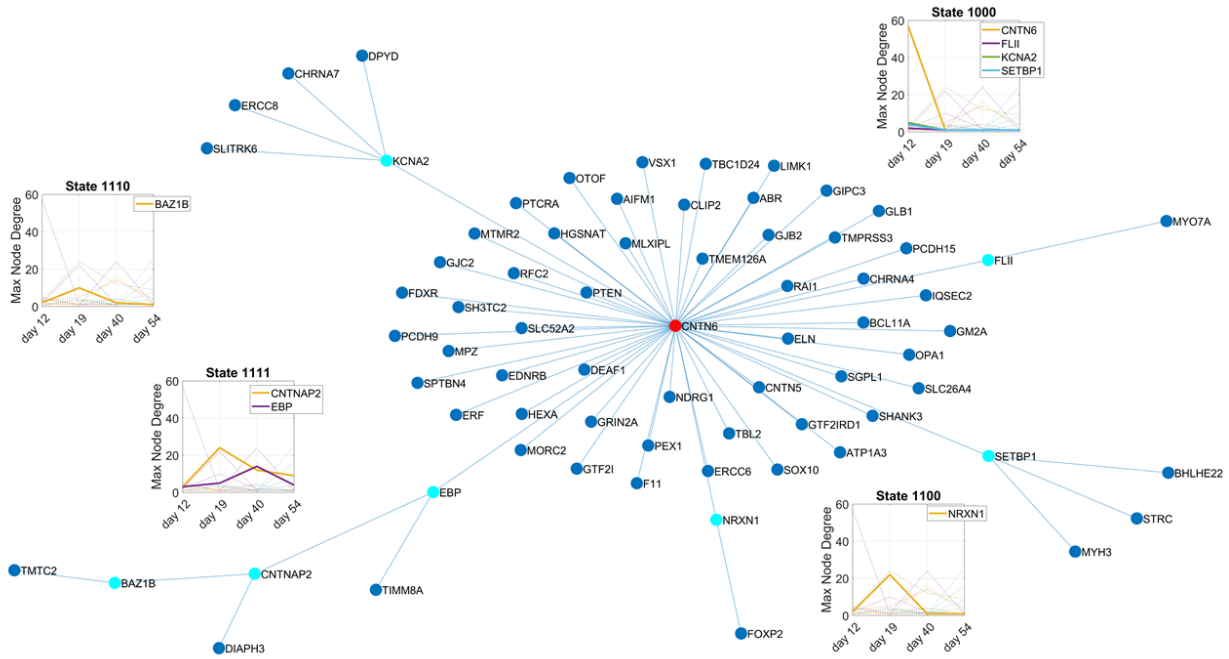


Figure 6 Probabilistic graphical model with maximum spanning tree, used to unveil variable degrees of co-dependencies in 72 genes shared between autism and brainstem disorders during early embryonic stages, as pluripotent embryonic stem cells become neurons. Main hub in day 12/54 days is CNTN6 (degree 57) identified in red and 7 sub-hubs (degree >1 but < 10) are identified with cyan color. Insets show the dynamic binary state paths, *e.g.*, of the hub CNTN6 and sub-hubs (FLII, KCNA2, SETBP1.) These genes expressions are ON (above the mean 1.9722 degree) during day 12 but turn OFF in subsequent days. Likewise, we highlight the paths for genes with states 1100, 1110 and 1111.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The application of the Chow-Liu Tree method on various sets of genes and their expressions, yielded that in the hESCs stages under consideration (days 12-54), most of the genes associated with brainstem malfunctioning, and shared with the autism associated SFARI genes, were leaf nodes (degree 1). Only a few genes had higher degrees and very few of them had significantly higher degree, *i.e.*, an order of magnitude above all others. This result suggests that the network degree could potentially be used to identify "key" genes for the various brainstem functions.

Table 1 summarizes the genes with maximal degree (overall maximum degree upon finding the maximum spanning tree, was 57). The genes with degree higher than 10 and those below 10 but above 1 degree are also featured. Supplementary Table 1 shows those with degree 1 (independence) for each day under consideration.

Day	Genes (Max Degree)	Max > Degree > 10	10 > Degree > 1
12	CNTN6(57)	none	BAZ1B(2), FLII(2), NRXN1(2), CNTNAP2(3), EBP(3), SETBP1(4), KCNA2(5)
19	CNTNAP2(24)	NRXN1(22)	DEAF1(2), HGSNAT(2), SHANK3(2), ABR(3), RI1(3), MTMR2(4), SGPL1(4), EBP(5),BAZ1B(10)
40	GTF2I(24)	CNTNAP2(12), EBP(14), ATP1A3(16)	BAZ1B(2), HEXA(2), PCDH15(4), RAI1(4),
54	GJB2(25)	GTF2IRD1(16)	OTOF(4), BCL11A(8), RFC2(8), CNTNAP2(9), GIPC3(4)

Table 1 Main hub and sub-hubs each measurement of the 12-59 days reveal key genes associated with brainstem and autism spectrum disorders.

Figure 6-9 show the evolution of the network's hubs and sub-hubs during days 12-54. These networks speak of the stochastic interdependencies derived from genes' expression. As such they speak of the genes' fate, as the cells become neurons. However, some genes may be downregulated or upregulated in different days, thus affecting their neighboring relations, and changing their state in asynchronous ways. To speak of the genes' state, we obtained the average degree each day. Despite the shifting hubs, each day the average degree was conserved, resulting in 1.9722. Then, the average degree overall (across the 4 days of measurement) was also this value and we set it as a threshold to denote the gene ON or OFF in a binary code. In four readings from 4 days, we obtained 16 binary vectors, 2^4. Then, we obtained the genes' state trajectory, reflecting both fate and state dynamics. Identifying the genes' binary states for those each day and forecasting each day according to future states, we then build the inset paths of each of the Figures 6-9. For example, in Figure 6, we see genes in state 1000 meaning that they are ON during day 12 but then turn OFF during days 19-54. On the other hand, we see genes that remain ON all days as 1111 (CNTNAP2 and EBP) both important for channels, synapses, and myelination processes.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Using this code, we can quantify the dynamic and stochastic evolution of the network's states and fate, which automatically identifies the hub and sub-hubs of the network each day. We can then interrogate the shortest distance paths from one gene to the other and uncover the hubs visited by these geodesic paths. In this sense, we explore the network's complex dynamics as the cells evolve from pluripotent to neuronal states. An example of such a path in day 54 is shown in Figure 9 (magenta), between SOX10 (22q13.1 involved in peripheral and central demyelinating neuropathy [36]) and FOXP2 (uniquely involved in human language, speech and orofacial dyspraxia [37-39].) This path alerts us of the importance of other hub genes (OTOF, IQSEC2, GJB2 (the main hub), REC2, GTF2IRD1, and EBP, connecting SOX10 with FOXP2. In this case, gene GJB2 provides instructions for making a gap junction protein (beta2, or conexin26) fundamental to form channel (gap junctions) that permit the transport of ions like potassium, fundamental in transduction processes. Connexin 26 is found in cells throughout the inner ear (important for the processing of sound) and the skin's outermost layer (important for the sense of touch.) This is just one path for the sake of discussion, but other paths could be explored to uncover the dynamics of genes' interactions as they evolve through early embryonic stages.

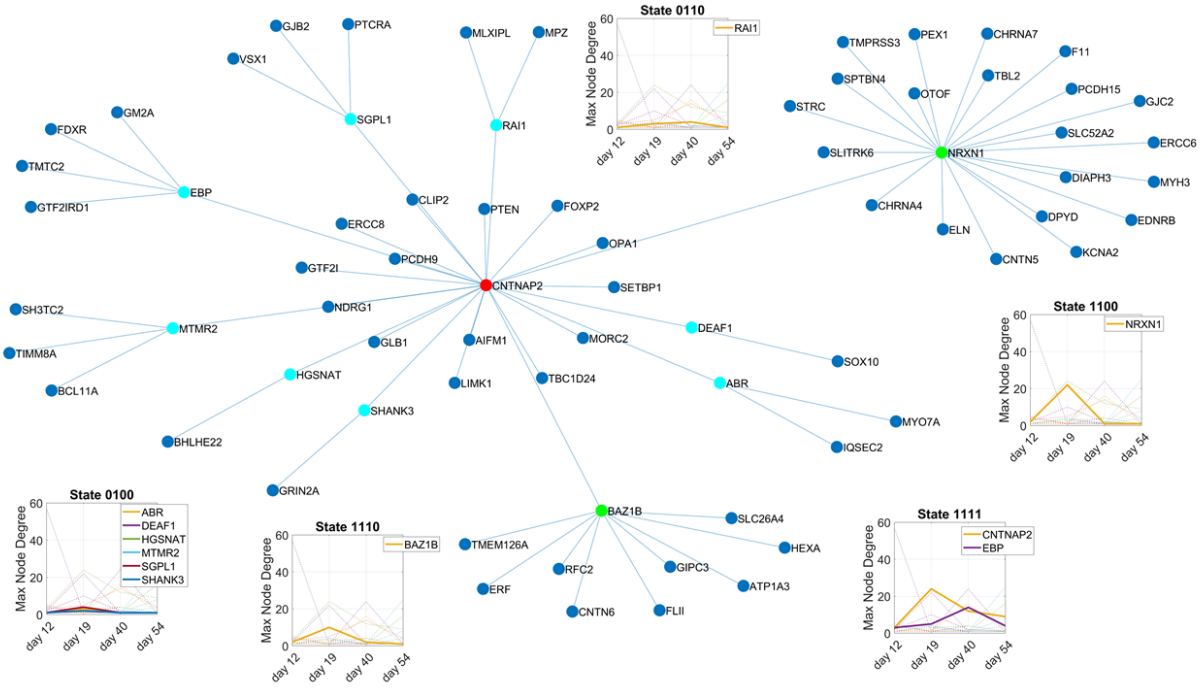


Figure 7 Network reorganization on day 19. The gene CNTNAP2 emerges as the main hub while new sub-hubs appear with degree ≥10 (green) and 10>degree>1. Genes with binary state path 0100 are ON in day 19 only, while the main hub is in the path 1111 which remains ON throughout. Other binary state paths are shown for the sub-hubs RI1 (0110), NRXN1 (1100) and BAZ1B (1110), all of which are OFF in day 54, when the ESCs are neurons but ON in intermediate days of early embryonic pluripotency.

Besides channel-level problems (*e.g.* GJB2), problems with demyelination (*e.g.* SOX10) and phenotypic rarities preceding absent spoken language (*e.g.*, not able to spontaneously cough or

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sneeze as a baby with FOXP2 deficiency), other genes in the network may contribute to increased delays in latencies at the molecular level, in the synapse, *e.g*. hub-gene NRXN1 (neurexin I) in location 2p16.3 is linked to severe epilepsy (Pitt-Hopkins-like syndrome), social affective behavior, and schizophrenia and autism [40-43]. We note that the affective neuropsychiatric disorders fall very close in our clustering results considering genes' expression across human tissues (Figure 5B). NRXN1 (code 1100) affects the synapse induction by neuroligins in early stages, and the generation of protein diversity [44]. Furthermore, neurexins induce differentiation of GABA and glutamate postsynaptic specialization via neuroligins [45], which are essential and have regulatory components at the synapse [46].

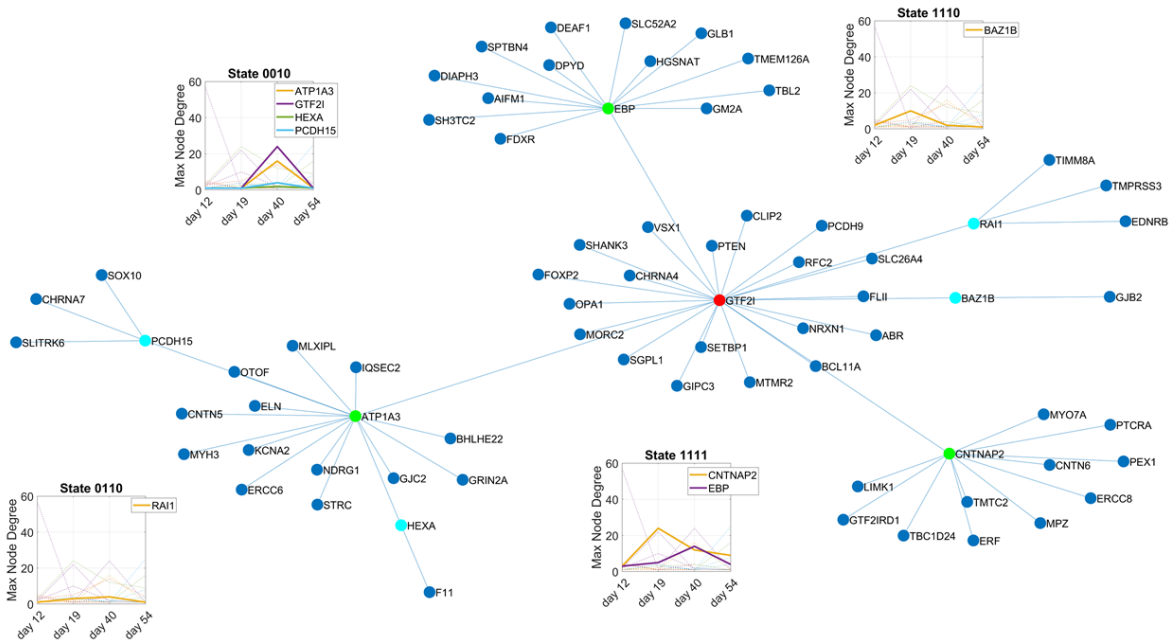


Figure 8 Network reorganization on day 40. Main hub (red) shifted to GTF2I (general transcription factor II-I in cytogenetic location 7q11.23. It turns on this day, but it is OFF all other days, 0010 as shown in the inset (with other sub-hubs of varying degree.) Other sub-hubs reorganize, demonstrating the asynchronous dynamics of the genes, and highlighting the importance of considering the full trajectory of the network of probabilistic co-dependencies.

Hub-gene CNTNAP2 (code 1111) plays a role in myelination patterns and signal transmission, needed for good communication across regions of the brainstem and between the brainstem and other systems. Specifically, this gene codes a neuronal transmembrane protein member of the neurexin superfamily involved in neural-glia interactions and clustering of potassium channels in myelinated axons. Rapid conduction in myelinated axons depends on the generation of specialized subcellular domains to which different sets of ion channels are localized. Other key genes to autism and disorders of the brainstem can be interrogated using these probabilistic co-dependencies, to better understand the dynamic and stochastic

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

contributions to the latency delays and the narrow bandwidth of information that we uncovered in the asd neonates' cohort.

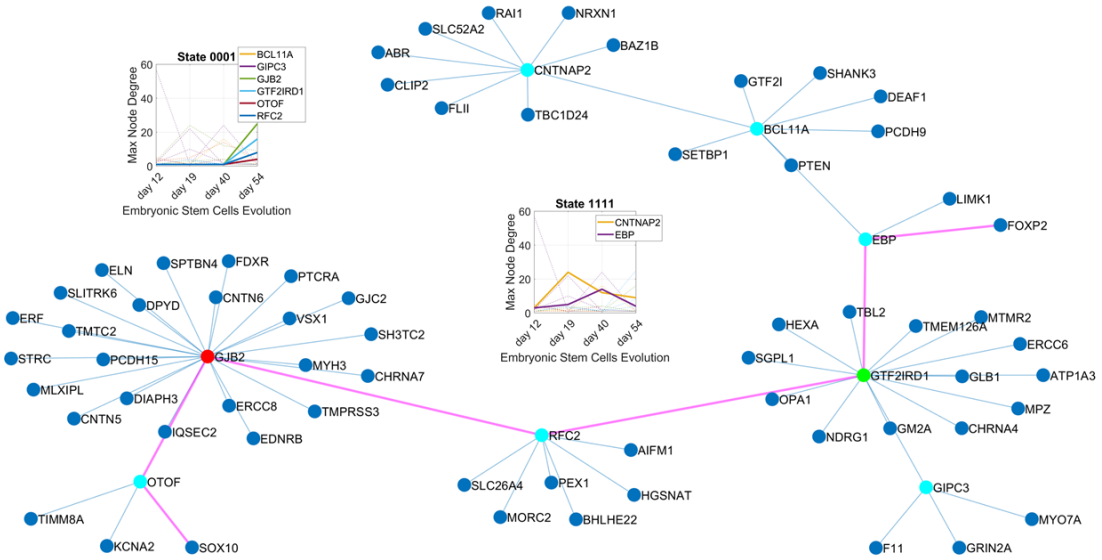


Figure 9 Reorganization of the network on day 54, during neuronal stage. Insets provide the main hub and sub-hubs which were OFF days 12-40 but ON day 54 (state path 0001) and those which remained ON throughout (state path 1111). A sample shortest distance path visiting the main hub and 4 additional sub-hubs is highlighted (magenta) between the SOX10 and FOXP2 genes (each being a leaf of this maximal spanning tree)

Discussion

The ABR constitutes one of the most reliable measures of neural integrity in the cochlear and caudal brainstem pathway, here indicating as well cumulative delays as the sound wave propagates and arrives at the primary auditory cortex in latency VII. This work, under the proposed Precision Autism paradigm [31], non-trivially expands our knowledge on ABR. We reveal specific information appropriate to infer the neurological and audiological status of healthy and special risk populations, such as the neonates in the autism group. We find that systematic shifts in latency lead to a delayed arrival of the sound at each point of interest along the auditory nerve, throughout the brainstem. Such cumulative delays were accompanied by significantly narrower range of latencies, suggesting narrower access to the bandwidth of frequencies experienced by the neonates. Furthermore, we identify important genomic networks and conditional co-dependencies that revealed key hub genes in early embryonic stages of neural development.

Under the unifying platform of Precision Autism, we examined stratification of the cohort as well as individualized profiles of behavioral performance. Furthermore, we tied this precision phenotyping to genomic data using two levels of analyses. At a higher level, we examined genes expression on 54 tissues and used marginal distributions to understand self-groupings across 21,033 genes and tissues ranked by expression levels. This revealed that genes associated

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with brainstem disorders and autism are closest in probability space of gene expression (on 54 tissues) to genes associated to CP and autism. This also ranked the cerebellum and cerebellar hemispheres as the tissue with the highest expressions of the relevant genes, followed by brain tissues involved in executive function, general motor control, emotion, navigation, and self-regulatory functions. At a lower level, we considered genes' codependent interactions as dynamically evolving networks during early embryonic stages of pluripotent embryonic stem cells turning into neurons. These analyses revealed hubs and sub-hubs as key genes and shortest distance paths interconnecting them and dynamically changing each day. Across 54 days and each day of 4 readings (on days 12, 19, 40, 54) we showed the fate and state trajectories of these main hubs and sub-hubs through a binary code. This code once again highlighted the importance of considering all fluctuations in gene expression as they impact neighboring co-dependent genes through asynchronous interactions during early embryonic changes leading to neuronal states. The results from the analyses of this compact set of genes tied to autism and brainstem disorders, warrants future dynamic interrogation of the full human transcriptome during different stages of early neural development.

An important corollary of the network analyses, particularly the interrogation revealing the shortest distance paths between pairs of genes involved at each level of the hierarchy of Figure 1A, is that we can see two problems that require our attention. The first problem is a many-to-one map between the levels of this hierarchy and the phenotype that they may give rise to. We may have a group of babies with the same phenotype, *i.e.,* expressing a common range of delayed latencies that nevertheless may have many different types of underlying disruptions at any given level, with different relevance at each level. Some babies may have disruption in genes primarily impacting the synapse level, others may have disruption in genes primarily impacting the circuits level, *etc*. Because of plasticity, coping mechanisms and epigenetic factors (among other factors), they may converge to a common phenotype. As such, from the top down, we see the many-to-one problem that needs to be resolved to develop true personalized targeted treatments. *Given a range of delayed latencies, find the levels that are most likely affected, if you know the transcriptomic data of the baby*. Dynamic interrogation of the transcriptome data as human embryonic stem cells evolve to neurons is underway (but further discussion is beyond the scope of this paper.) The many-to-one problem has (at least) a known general mathematical solution that could produce a locally linear isometric embedding of the lower dimensional phenotype space (of I-VII latencies)) into the larger dimensional space of hierarchical levels (with driving genes) [1]. Therefore, given the latencies, proper genomic information, and some knowledge about environmental and clinical family history, one can attempt to approximate the levels and genes to target for the development of treatment, to help the person. Furthermore, the phylogenetic order of these levels contributes to naturally reducing the problem to detect hubs of high importance at the very basic molecular, channels and synaptic levels, impacting all other levels above. This orderly nesting leads to a priority map, to prioritize survival over daily function, daily function over abstract cognition, *etc.* in relation to intactness of fundamental tissues such as cardiac, smooth, and skeletal muscle types, and brain-body regions.

The second problem emerges from the bottom up. There are one-to-many maps spanned at each level and across levels. This is so, because at each level, there are cascade effects impacting and

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

percolating across the next levels above. Inevitably, this will give rise to a multitude of phenotypes with different ranges of delayed latencies at each of the I-VII sites under consideration. We will need to further refine the contributions of genes impacting each level to the level above and interacting as a network with conditional nodes' dependencies that include causal directionality. For example, if the baby has problems with genes impacting gap junctions, we will need to go beyond the gap junction level and assess how much each of the subsequent levels above the molecules and synapses levels are impacted and driving other subsequent levels above theirs. It will be necessary to go back and forth from the precision biometrics (digital phenotyping) to the levels where we are forced to make inferences -perhaps due to lack of access to data at that level. For example, we may get access to neuronal classes and infer circuits and systems via iPSCs tracking, then record circuits and systems level data in ways that are far from realistic (because that is what we can do with current instruments/models.) The one-to-many problem is particularly interesting because of the cascade effect percolating upwards and impacting each level with variable degree for each person. Considering both the top-down and bottom-up approaches together, in a non-reductionist way will be critical to advance the field. Thus, collaborative thinking and multidisciplinary cooperation among researchers specializing on the knowledge base of each level, will be critical in the next decades of research. In this sense the recently proposed Precision Autism paradigm [31] offers a research program to help advance these questions that need to be resolved at each level of inquiry in Figure 1A.

Important prior work had proposed ABR signals as an early biomarker flagging neurodevelopmental issues [20, 29, 47, 48] but no connection to the genomic landscape had been provided. The present work offers a new unifying platform amenable to connect multiple layers of the knowledge network (as in Precision Medicine [49]) adapted to autism and other neurodevelopmental disorders currently falling under the autism umbrella term [31]. Indeed, we combined clinical reports with ABR at the micro-level of behavioral precision, using fluctuations in the waveforms' amplitude and peaks' latencies. Such data would have been otherwise smoothed out as superfluous noise (gross data) under traditional approaches. We further leveraged known genomic information linked to both autism and brainstem-related disorders, in the context of 14 neuropsychiatric and neurological disorders, to situate early brainstem disorders in relation to disorders that are diagnosed later in infancy, childhood or young adulthood (*e.g.*, CP, autism, ADHD, schizophrenia and depression.) There, CP was identified as the most similar in a probability space of genes' expression on human tissue. Our new integrative approach was then extended to identify subtypes of the broad spectrum of neurodevelopmental disorders under the autism label, to stratify groups for treatments' design (*e.g.*, as in Figure 5). Examining these identifiable subgroups, in future studies, we could further refine genomics and proteomics knowledge to include network interactions and emerging hub genes that evolve along differentiable trajectories in the early embryonic stages. Given the person's induced pluripotent stem cells iPSCs (*e.g.*, derived from skin or blood samples), this approach would enable true personalized search of targets for treatments, tracking *the full transcriptome* in the iPSCs cells as they transition from pluripotent to neuronal states.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Leveraging the plasticity of the nascent nervous systems, the proposed platform could also help us identify physiologically based behavioral targets for treatment. At the level of ABR outcomes, we see the need to aim for a broader range of latencies in the neonate's responses, *i.e.*, broaden their response frequencies to sound levels, and shift the centroid of the probability distributions of latencies to coincide with those of neurotypical ranges. Gradually shifting the delays in latencies at each accessible level (I-VII) of the neonates with signatures like those who went on to receive the asd diagnosis, would ensure alignment between the endogenous sensory experience of the baby and the exogenous sensory experience generated by others of his/her own kind. This alignment would help creating a coherent percept of the world for the child to be able to interact with another human along similar timescales and temporal coincidences.

At the neural systems level, and at the level of synapses and channels, we will have to uncover and pursue other targets for treatment, to achieve proper signal transduction and transmission latencies across the brainstem leading to proper coincidences for effective communication with an interlocutor. This work under the Precision Autism model [31], offers a new avenue to integrate knowledge across layers spanning from molecules to complex social behaviors. As discussed above, further refinement is possible at the levels of molecules, channels, myelination, using network of genes dynamically interrelated through conditional dependencies. As with the evolving human transcriptome, these new methods open many new lines of inquiry to develop dynamic models of full transcriptome interrogation, considering also gross data from asynchronous changes over time. During early embryonic stages, such changes are bound to impact the shifts in network hubs as cells transition to neuronal states and circuits start to develop into systems of networks and sensory-motor maps.

Given their phylogenetic orderly maturation in humans, perhaps addressing the brainstem issues earlier would help better support the later development of somatic sensory motor systems. Along these lines, acousticomotor reflex reactions in human infants provide a window into the integration of sensory information across systems undergoing disparate maturation rates. Those systems receiving endogenously and spontaneously self-generated input from the periphery must integrate and inversely transform exogenous information centrally processed, such as sound and light, into sequences of synergistic movements across the face, the neck (for speech and swallowing), eyes, and body. The success of such operations may be key to form spatio-temporal bodily maps for effective actions and interactions with others. It would be impossible to mirror social behavior or to understand it if the incoming signals are sensed as echoes. No simultaneous percept of the body in space and time could then be attained for the person to understand his/her spatio-temporal self-existence, let alone that of others in the social scene. It would be very difficult to relate to others under such conditions. Clearly, it would be difficult to relate to others, not because the person did not want to, but rather because the person's brain-body function is operating (literally) on a different spatio-temporal scale and having access to incomplete frequency ranges of the electromagnetic spectrum accessed by the various sensory organs of the nervous system.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

At the crux of this complex matter is a well-known coordinate transformation problem that the brain needs to resolve in order to control the excess degrees of freedom that enables the generation and maintenance of smooth behaviors [1], while managing the stochastic variations of those DOF differently, according to levels of intent or spontaneity [17]. This coordinate transformation links the vast endogenous world of the person with the exogeneous range of sensory inputs in ways that allow, from the start of life, the transition from exploratory to goal-directed (error-correcting) codes [50]. Such codes are bound to emerge through a process of self-exploration and self-discovery that builds spatio-temporal maps of actions and their possible sensory consequences. Eventually they distinguish systematic, predictive codes of intentional movements from spontaneous random patterns in writhing or fidgeting movements. Without the ability to transform sensory to motor goals in ways that ultimately anticipate and compensate for such temporal delays, we would be forever trapped in random noise, lost in echoes of ourselves and others in the world, trying aimlessly to find that anchor that enables the simultaneous experience of us in the world and of the world in us.

Evidence from ABR studies suggests that prenatal maturation in the central auditory system [15, 51] far precedes sensory motor maturation. As such, when excess delays in the latencies of signal propagation along the brainstem structures are detected in the newborn, concerns about proper neurodevelopment emerge [47]. It is in this sense, that as of today, neonatal delays in ABR could constitute the earliest digital biomarker for neurodevelopmental derail, detecting the problem even earlier than the recently proposed motor-sensing signatures [7, 52-54]. The latter (within the first trimester of life) far precedes those unambiguously visible differences in social interactions. Current inventories rely on observation of such unambiguous differences in social interactions that surface much later, after two or three years of age [55]. Keeping this phylogenetically orderly timeline in mind is important for early detection of departures from neurotypical development. Nevertheless, current one-size-fits-all early intervention programs are not designed to differentially address each level of maturation and tailor the program according to each neonate subtype.

Caveats and Limitations

This is a retrospective data set, which was collected over the span of 17 years (2001-2018.) Although the recording protocol remained the same, the ABR equipment was upgraded twice within this period, with efforts made to calibrate the sound outputs of the newer equipment to match the older one. Given that these are clinical recordings of neonates, obtained under practical, realistic situations, inevitably, several recordings were noisy, which had to be excluded from the analyses. The data in fact came from an opportunistic sample designed without any prospective analytics in mind. The ages of diagnoses varied due to the different focuses of the project based on grant fundings. It would probably have been better if the dataset came from a more compact study with more consistent form of diagnoses. The results, which stand despite these caveats, nonetheless warrant further advancing such study with a larger cohort.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Conclusion

Neurodevelopmental asynchronous differences, owing to early derail in this orchestration of disparate endogenous temporal dynamics from systems with different maturation levels at birth, can result in diverse coping strategies for the affected babies. More generally, these coping strategies could contribute to highly heterogeneous ways to socialize, eventually turning into observable aspects of complex social behaviors [56, 57], far easier to detect than those hidden to the naked eye of the observer [58-60]. Because of their early manifestation, clinical availability and technical rigor of its detection, general shifts in response delays from ABR latencies in infancy, may offer today the most robust way to address the problem of early neurodevelopmental derail and further offer ways to individualize early treatments.

Given the hubs detected in early embryonic states and the fate and state codes developed here, we appreciate key genomic contributions to disparate delays at different levels. These span from molecules to systems and interconnected cortical, subcortical, and cerebellar networks that seem crucial to develop proper maps and systems across the brain and body. As such, we suggest here that the brainstem is well positioned to serve as a sort of *clock* referencing endogenous delays along different somatic-sensory-motor time scales and aligning them onto a common anchor. This anchor would then enable experiencing the world simultaneously as a coherent percept. Without it, autism, as we know it today in its broad spectrum, will inevitably emerge.

Methods

Participants

Infants in this study are a subset of participants from our longitudinal study of attention and arousal in children at-risk for developmental problems due to pre-term birth and/or medical factors requiring assignment to the NICU [61, 62]. Selection criteria for the larger study included any of the following: birthweight (BW) (<1,800 g); fetal distress with evidence of asphyxia at birth; assisted ventilation (>48 h); persistent apnea or bradycardia; abnormal neurological signs; small for gestational age (<10th percentile BW for gestational age), intrauterine growth restriction, or dysmature; multiple gestation (at least one twin met criteria or BW <2,000 g). Exclusion criteria were known prenatal exposure to drugs of abuse, diagnosis of a major congenital anomaly, or chromosomal disorders. In addition, a sample of healthy newborns, recruited from the newborn nursery, participated in the original study as controls. Additional criteria for the current study were:

1. Administered an ABR during the newborn period.
2. Assessed at older ages to ascertain or rule out an ASD diagnosis.
3. Absence of significant brain damage documented by cranial ultrasound, CT, MRI such as IVH (Papile Grade >I), ventriculomegaly (>5 mm), periventricular leukomalacia, hydrocephalus, or seizures requiring treatment documented by EEG.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The data that we analyzed pertained to a subset of children from this study, with *n=24* receiving an asd diagnosis (18 NICU; 6 WNB). Diagnoses were determined as follows: by clinicians at the Institute for Basic Research in Developmental Disabilities (*n=9*); through participation in Early Intervention or special preschool programs (*n=8*); by parent interview with the ADI-R (*n=6*); by a private neurologist (*N=1*). A control group comprised of 30 children was selected (19 NICU; 11 WNB). No child in the control group was diagnosed with autism or was suspected of being on the autism spectrum, based upon behavioral testing at their 24/36-month follow-up assessments. In addition, based upon the PDD Behavior Inventory[63], none were reported to show any signs of asd by their mothers; none were receiving special services because of asd or suspected asd.

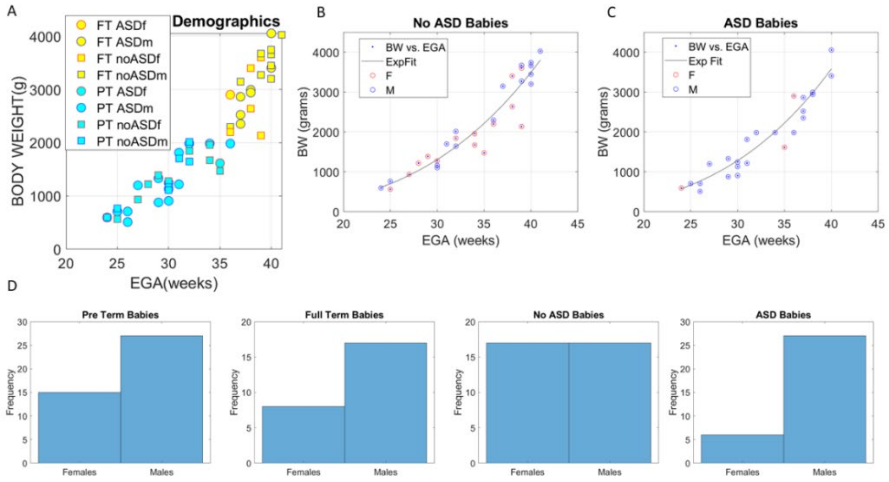


Figure 10 Demographics of the study. Body weight (BW) expressed as a function of estimated gestational age (EGA), color coded as full-term (yellow FT) or pre-term (cyan PT) categories. Marker edge denotes sex (red female, blue male) and marker type denotes diagnosis type (circle asd, square non-asd). (B-C) Nonlinear relation between estimated gestational age (EGA) and body weight (BW). (D) Available *trials* from babies with complete demographic records available per baby type used in supplementary material figures featuring population statistics.

Participants comprised 64.9% males/35.6% females and were ethnically diverse. They were 31.2% Latinx (including Latinx black, white, Asian, and mixed) and 68.8% non-Latinx, with 16.9% black; 54.5% white; 3.9% Asian, 1.3% Indian, and 23.4% mixed. The research protocol was approved by the Institutional Review Boards of involved institutions and written informed consents were obtained from parents/guardians of all participants.

Auditory Brainstem Response (ABR)—ABRs were administered during the newborn period. For NICU infants they typically occurred at the bedside shortly after birth, but no earlier than 24 hours postnatally and 31–32 weeks post-conceptional age. (Mean (SD) age at test was 37.5 (4.1) weeks). Ipsilateral left ear vertex-to-mastoid recordings were made using 100 μsec rectangular rarefaction click stimuli 80 dB above adult normal hearing level (nHL) presented at a rate of 12.9/s. Our standard average ABR waveform consisted of 3,072 (3 sets of 1,024) artifact-free trials. Two additional sets of 2,048 (2 sets of 1,024) trials were recorded at 75 and 70 dB nHL. Stimulus delivery and response averaging were done using Intelligent Hearing System Brainstem Smart EP System.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Clinical Tests: In the cohort prior to 2011, beginning at 3 months of age, and extending through 25 months, infants were tested with the Bayley Scales of Infant Development - 2nd Edition (BSID-II) (Bayley, 1993). The BSID-II is an individually administered scaled neuro-developmental assessment that measures cognition, motor, and behavioral development. It has been used widely both in research and clinical applications and is currently in its 4th edition. It comprises Mental, Motor, and Behavior Rating Scales. Only the first two scales were analyzed in the current study. Age-appropriate norms were applied to the scaled scores to derive a Mental Developmental Index (MDI) and a Psychomotor Developmental Index (PDI). Both indices have a mean of 100 and a standard deviation of 16.

Infants in the cohort after 2011 were assessed from 3- through 24-months of age with the Mullen Scales of Early Learning (MSEL AGS Edition, 1995). The MSEL is a comprehensive measure of cognitive and motor functioning divided into five areas including Gross Motor, Visual Reception, Fine Motor, Receptive and Expressive Language. It yields separate scores in each of these domains as well as a composite Standard Score made up 4 of the 5 areas (Gross Motor is excluded).

Micromovement Analyses

Figure 11 explains the derivation of the micro-movement spikes, a general datatype that standardizes the timeseries waveforms for further analyses that involves diverse populations, while preserving the original ranges of the data. Micro-movement spikes turn continuous analogue signals to digital spike representations and permit analyses of the noise to signal ratios of the time series data and of their information content. They have been successfully extracted and characterized in time series of resting state functional magnetic resonance imaging data, kinematics data from positional sensors and inertial measurement units, electroencephalogram, electrocardiogram, and electromyogram. They have also been used with success to stratify disorders of the nervous systems using timeseries of genomic data. Because these micro-movements come from data that is traditionally smoothed out as noise (gross data) in current practices, they tend to non-obviously reveal information in well-study phenomena, that has otherwise remained hidden up to now. The notion of micro-movements has been patented by the US Patent office [64]. Here we combine the resting state jitter preceding the Dirac-delta like pulse and the jitter from the post-pulse refractory period with the pulse, to then extract the micro-movement spikes from the full sequence of ABR points collected at 25kHz.

For Peer Review

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

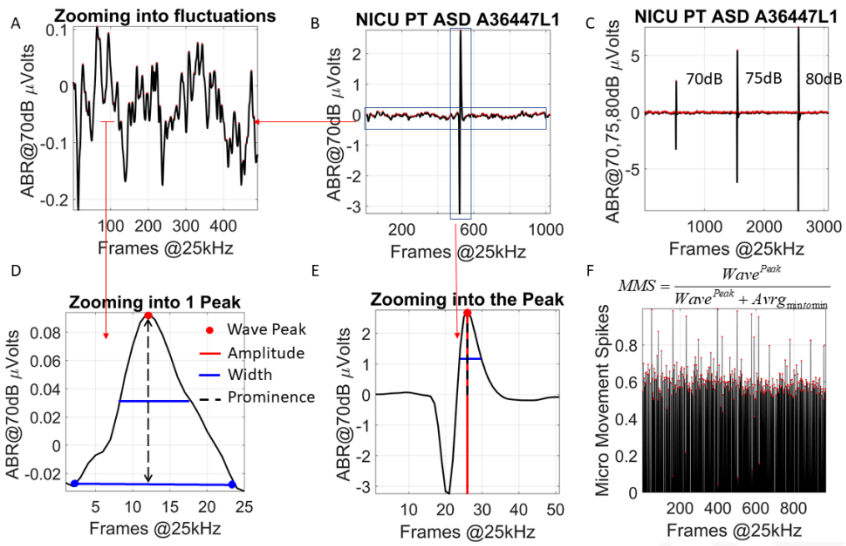


Figure 11 Derivation of the micro-movement spikes. (A-B) Raw waveform in (B) represents the auditory brainstem evoked response ABS from one click at *e.g.* 70dB. (C) Sample ABRs from three clicks at 70, 75 and 80dB. (D) Each pulse (Dirac delta) is preceded and followed by microscopic jitter that are zoomed in panel (A). The full waveform is thus obtained, and the amplitudes (μV) and latency of the peaks (jitter and Dirac delta combined) used to scale as in panel D, where we show the peak amplitude, width, and prominence features under consideration. These features are extracted from both the jitter and pulse. (E) Zooming into the pulse to highlight the features under consideration in (D). (F) Resulting micro-movement spikes derived from the jitter and pulse features (in this case for the prominences) using the formula in the panel to scale out allometric effects due to anatomical differences that emerge from different body sizes and masses rapidly changing in early infancy. The micro-movements data type can be derived from any timeseries by obtaining the absolute deviations from the mean amplitude or inter-peak-interval timings, or any other feature of the waveform that takes then the series of fluctuations away from the mean and scales them appropriately between 0-1. This standardized waveform can then be used in further analytical pipelines to compare participants of disparate ages and anatomical types.

Stochastic Analyses

Figure 12 explains the estimation procedure and shows the analytical pipeline to derive interpretable parameter spaces and make empirically informed statistical inferences from the data. This method enables optimal feature identification and parameter spaces to stratify phenomena in general and uncover important patterns in the data. Similarity metric in Figure 13 (Earth Mover's Distance, EMD [65, 66]) can be used to measure distances between points in probability space and investigate the rates of stochastic shift.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

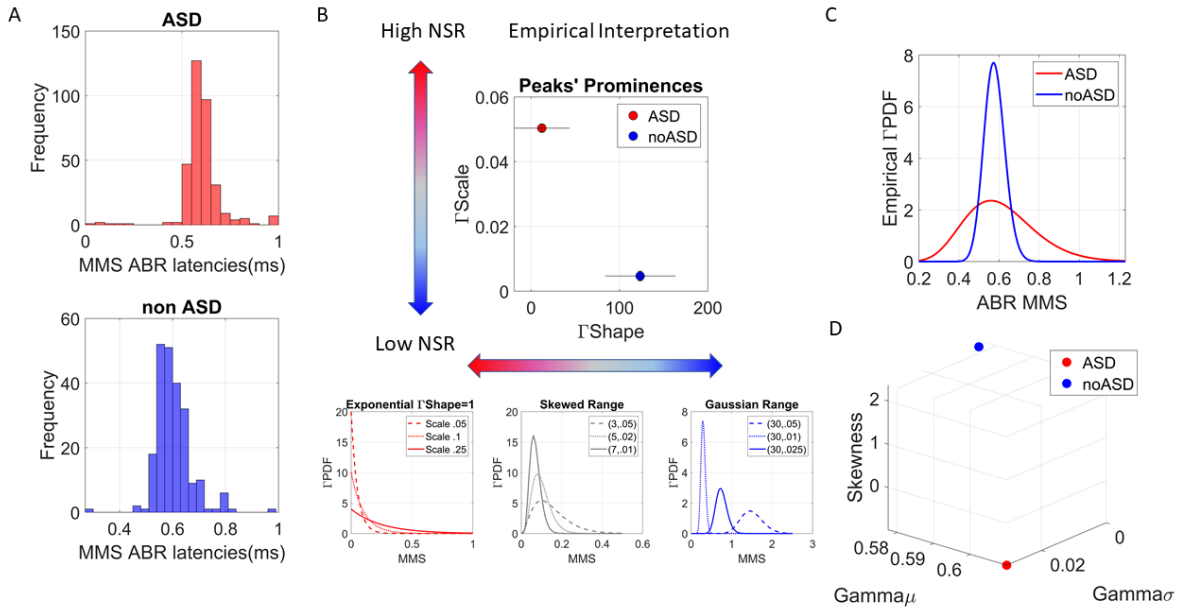


Figure 12 Empirical estimation and inference of stochastic signatures derived from the micro-movement spikes. (A) Sample frequency histograms of micro-movement spikes derived from the fluctuations in the peaks' latencies (*ms*) normalized to be unitless as in Figure 12. (B) The Gamma parameter plane spanned by the shape and scale parameters empirically estimated using maximum likelihood estimation (MLE). The estimated parameters are plot with 95% confidence intervals for each estimated shape and scale value. These two data points correspond to the PDFs in (C). Gradient colored arrows denote the empirical ranges of human micro-movements PDFs, as previously obtained from different biosensors across different systems and levels of control, in several thousand human participants across the lifespan. Notice that because the MMS are standardized and scale out anatomical disparities, we can place points from different participants with different anatomical features (due to different ages) on the same parameter space. Along the shape dimension, values range from 1 (Exponential distribution representing a memoryless random process and being a special case of the continuous Gamma family) to over 100, for the Gaussian distribution-like symmetric case, with skewed distributions falling in between these two extreme cases. Along the scale (dispersion) dimension, lower values represent lower noise to signal ratio (NSR). This is so because given the Gamma variance and mean, and given the Gamma shape (a) and the Gamma scale (b), $NSR = \frac{a \cdot b^2}{a \cdot b} = b$ (D) The empirically estimated Gamma moments can be obtained as well and represented on a parameter space spanned by the mean (x-axis), the variance (y-axis), the skewness (z-axis) and the kurtosis proportional to the size of the marker. Lower NSR (blue non-asd) has lower variance in this case. Given the empirically estimated MMS values, we can make well informed inferences with highly interpretable value about the stochastic signatures. This is so because we have characterized them in humans before.

Genetic Analyses

DisGeNET genes: These included genes associated to auditory neuropathy spectrum disorder, auditory neuropathy, abnormal auditory evoked potential, central auditory processing disorder, auditory perceptual disorders, absent brainstem auditory responses, auditory inattention, auditory neuropathy, autosomal dominant 1, prolonged brainstem auditory evoked potentials, auditory neuropathy and optic atrophy, auditory neuropathy non-syndromic recessive and hyperacusis.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SFARI genes: These included 906 genes associated to autism spectrum disorders and ranked by significance according to the literature as scores 1-4. The SFARI Autism categories that I used to determine the level of confidence that the gene is linked to Autism, were those reported as of 03-04-2020.

GTEx: We used the Genotype-Tissue Expression (GTEx) project involving human RNA-seq, expressed in Transcripts Count Per Million (TCM) to examine the genes' expression across the 54 tissues sampled in their database [32]. Using this atlas of genetic regulatory effects across human tissues, as previously published work [31], I compare across diseases, for those genes overlapping with the SFARI Autism-linked genes, the common tissues where the expression of these genes is maximal.

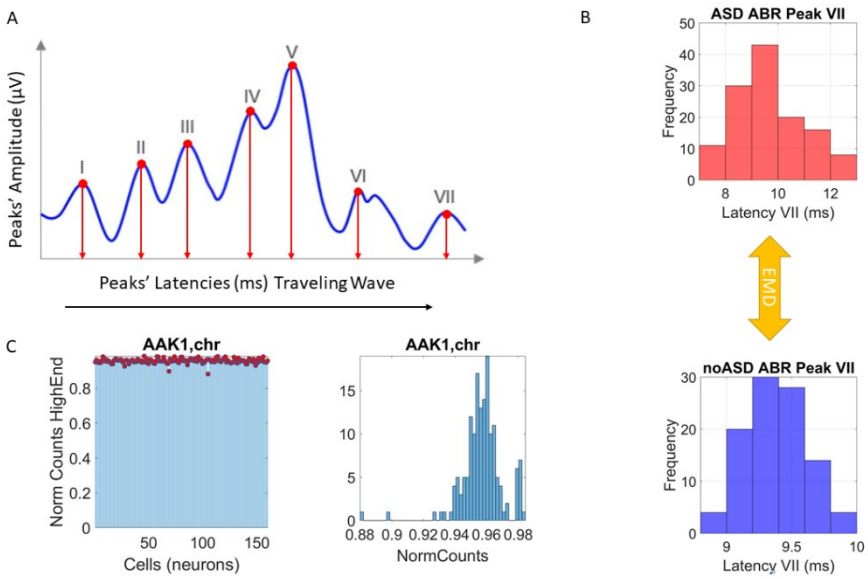


Figure 13 Similarity metric to determine pairwise distances between probability distributions. (A) Sample (schematic) waveform with standard peaks corresponding to sites I-VII. Latencies of the traveling wave are then recorded for each site and data gathered across the same number of trials for each baby (personalized approach) or for each cohort (*e.g.*, asd *vs.* non-asd, or pre-term *vs.* full-term.) (B) Sample frequency histograms of trials for site VII to compare asd *vs.* non-asd newborns using the Earth Mover's Distance providing a notion of similarity between the two distributions. Then a matrix of pairwise EMD values properly normalized between 0-1 can be obtained, whereby each entry of the matrix represents the pairwise distance (as in Supplementary Figure 2) between two groups, two babies, *etc.* (C) This same approach is used to determine distances in the space of genes' expression and obtain the normalized matrix of EMDs (*e.g.*, Figure 5A expressed as a colormap.)

Transcriptome Interrogation

Single-cell RNA-seq during neural differentiation of hESCs reveal a multitude of neural cell classes with a range of early brain regional identities in [34]. They analyzed 2,684 cells with >20,000 transcripts as assessed by unique molecular identifiers (UMIs). Cell types are named by point of origin as progenitor (P), transitional (T), neuronal (N) or other tissue (O). We focus on the evolution of the 26,046 genes' expression across the neuronal type examined at days 12, 19, 40 and 54. These data are public access and was downloaded and curated to apply our methods.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

We have tables with rows of 100s of cells by columns of genes' expression in transcriptome count per millions (log normalized) and we represent them as matrices. We then transpose the matrix and express each gene as a series of expressions across the number of cells (263, 168, 346 and 495 cells in days 12, 19, 40, 54 respectively.) Each day, we normalized the counts (0-1) values and for each gene (with equal number of cells) we obtain the frequency histogram and marginal probabilities. We then use the Earth Mover's Distance metric to quantify pairwise similarities between genes and build an adjacency matrix for further interrogation (beyond the scope of this paper.) The cumulative EMD per gene gives a window into the trajectory of the gene's expression to fate on day 54, while its state expressed as a binary value (ON or OFF) each day can be obtained as well (see below the graphical analyses to determine the degree of each gene as a contributing node to the network of conditional dependencies) to track the evolution of both fate and state of the whole dynamically changing transcriptome as the hESCs become neurons. This contrasts with traditional approaches of eliminating odd patterns of variations or asynchronous fluctuations in individual genes that may be transiently turned OFF and come back ON in later stages of differentiation. We approach the genomic information through the lens of complex, nonlinear stochastic dynamics in interacting networks where neighboring codependences mark important gene's hubs and sub-hubs linked across relevant (optimal) geodesic paths on the network.

Network Analyses

Undirected graphical modelling for evaluating the dependency between gene expressions

Graphical models have been extensively researched and used to describe statistical dependencies between random variables. These models can be either undirected graphs or directed graphs, in the latter case we can derive cause-effect relationships between the variables. In the current project we are interested in undirected graphical models.

Formally, for any set of random variables $X = (X_1, X_2, \dots, X_N)$ a graphical model attempts to associate the joint random vector X drawn from the probability space $(\Omega_1 \times \Omega_2 \times \dots \times \Omega_N, F_1 \times F_2 \times \dots \times F_N,\ P_X)$ with a graph $G = (V, E)$, where V stands for vertices and E for edges. Here, $\Omega_1, \Omega_2, \dots, \Omega_N$ are the sample spaces for each random variable and $\Omega_1 \times \Omega_2 \times \dots \Omega_N$ the joint space of the random vector X. $F_1, F_2, \dots, F_N$ are the corresponding generated Borel σ-fields to denote the sets of all possible random events for each random variable and P_X is the probability measure on the random vector X. The set of nodes V represents the random variables $X_1, X_2, \dots, X_N$ drawn from the probability spaces $(\Omega_1, F_1,\ P_{X_1})$, $(\Omega_2, F_2,\ P_{X_2})$, … $(\Omega_N, F_N,\ P_{X_N})$, with their respective Borel σ-fields and probability measures. An edge $e_{ij} \in E$ if and only if the random variables X_i and X_j depend on each other.

If two nodes u_i and u_j are not connected with an edge it implies that the two variables X_i and X_j are conditionally independent, *i.e.*, statistically independent *given* all other nodes:

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$X_i \perp X_j \mid X_{V/\{u_i,u_j\}}$$

This property of the graphical model is known as the *global Markov property*.

Estimation of a graphical model using Chow-Liu Trees

If we want to factorize the joint probability distribution in a dependency graph that has a tree structure, *i.e.*, every two nodes are connected by no more than one path (in other words there are no loops in the graph), then the joint density of the random vector X factorizes with respect to the pair-wise joint and marginal densities as:

$$f(x) = \prod_{e_{ij}\in E} \frac{f_{ij}(X_i, X_j)}{f_i(X_i) f_j(X_j)} \prod_{u_i \in E} f_i(x_i)$$

It turns out that in the case in which all variables are categorical and take values from a finite set it is very easy to find the optimal tree that factorizes the joint distribution. Let N_x be the number of times a realization x of the random vector X appears in a collection of independent and identically distributed (*i.i.d.*) samples. The tree G that optimally factorizes X, given the sample data Z of size n, will be the one with the maximum log-likelihood (MLE):

$$G = argmax\ L(G) = \sum_{x\in Z} N_x \log\left(\hat{f}_G(x)\right)$$

Which turns out to be:

$$L(G) = \mathrm{n} \sum_{e_{ij}\in E} I\left(\hat{f}_{ij}\left(X_i, X_j\right)\right) + \mathrm{C}$$

Where C is a constant and $I\left(\hat{f}_{ij}\left(X_i, X_j\right)\right)$ the empirical mutual information between X_i and X_j. Therefore, by choosing the appropriate subgraph G that maximizes the sum of the empirical mutual information estimates, we obtain the optimal tree structured graphical model. Since the model is a tree, we simply need to find the Maximum Spanning Tree from the mutual information network, *e.g.*, by using Kruskal's algorithm. The process we just described is known as the Chow-Liu algorithm and the extracted conditional dependency tree that factorizes X is the Chow-Liu tree.

Graphical Model and Maximal spanning (Chow-Liu) Trees of Genes' Co-Dependencies

For a specific tissue or organ, in our case the Brainstem, we have a set of genes associated with it. On a particular day D of cell maturation, we have the data $X^D_{N\times M}$, where N is the number of genes and M the number of cells. Each row of our data refers to a specific gene and each column to a specific cell. We treat each column (corresponding to a cell) as an *i.i.d.* sample drawn from the joint probability space of the expressions of that set of genes and we want to generate the undirected graphical networks $G_{12}, G_{19}, G_{40}, G_{54}$ for days 12, 19, 40 and 54.

In the case of continuous variables, the methods used to estimate the Chow-Liu tree usually involves Kernel Density Estimation (KDE) of the joint and marginal probability densities [67]. However, in our case we want to factorize the joint probability density of a gene expression

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

network with number of genes in the order of magnitude ~10^2. The application of KDE would require in this case, number of samples far exceeding the available number of ESCs. Therefore, we resort to extracting the Chow-Liu Trees by estimating the mutual information through binning and histograms on the available data. Once the Chow-Liu Tree corresponding to the factorization of the gene set of interest is obtained, we order the nodes in ascending degree. The proposed methodology can be appreciated in Figure 14 of the methods section.

What do we achieve with this ordering? It is obvious that the higher the degree of a gene is the more co-dependent his expression is with the expression of other genes in the network. This implies, in a statistical sense, that mutation or deletion of this gene is bound to immediately effect the expressions of many other genes. Note that this rationale simply states the co-dependency between the gene expression but the actual mechanisms through which this statistical relationship takes place are open to investigation.

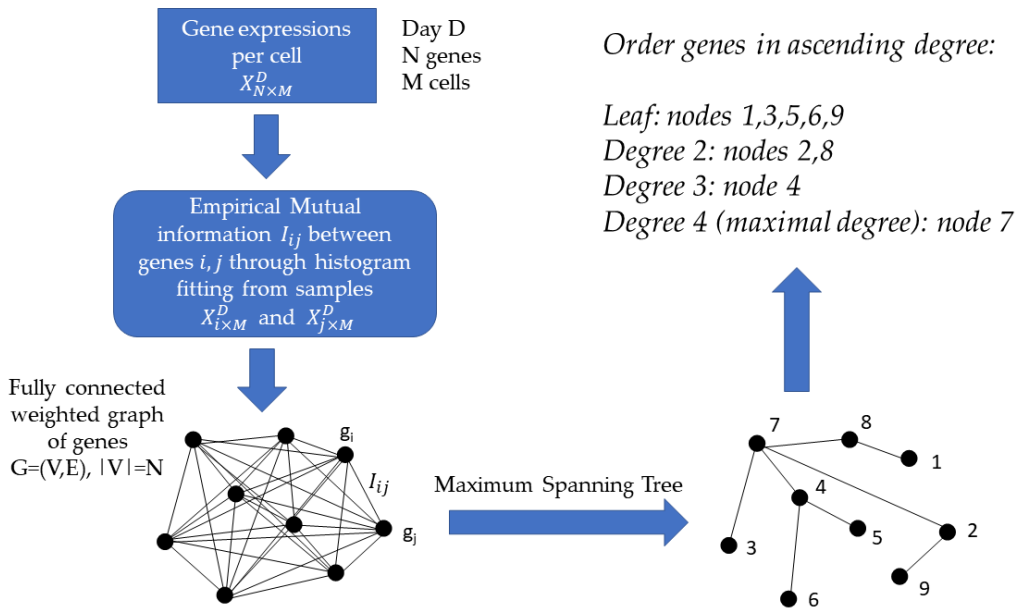


Figure 14: Proposed pipeline for the factorization of the joint probabilistic behavior of a network of genes and for determining the significance of each gene in the network

The research data collection was supported with funds from: New York State Office of People with Developmental Disabilities (OPWDD) and New York State Institute for Basic Research in Developmental Disabilities (IBR); NICHD # P01-HD047281 and Autism Speaks # 7598 (Gardner). EBT and TB were partly supported by an endowment from the Early Career Development Award to EBT (2014-2017) by the Nancy Lurie Marks Family Foundation.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Torres, E.B., *Theoretical Framework for the Study of Sensori-motor Integration.*, in *Cognitive Science*. 2001, University of California, San Diego: La Jolla. p. 115.
2. Torres, E.B., Zipser, D., *Reaching to Grasp with a Multi-jointed Arm (I): A Computational Model.* Journal of Neurophysiology, 2002. **88**: p. 1-13.
3. Torres, E. and R. Andersen, *Space-time separation during obstacle-avoidance learning in monkeys.* J Neurophysiol, 2006. **96**(5): p. 2613-32.
4. Yin, H.H., *The Basal Ganglia in Action.* Neuroscientist, 2017. **23**(3): p. 299-313.
5. Yin, H.H., *Action, time and the basal ganglia.* Philos Trans R Soc Lond B Biol Sci, 2014. **369**(1637): p. 20120473.
6. Torres, E.B., *Objective biometric methods for the diagnosis and treatment of nervous system disorders*. 2018, London: Academic Press. x, 568 pages.
7. Torres, E.B., et al., *Neonatal Diagnostics: Toward Dynamic Growth Charts of Neuromotor Control.* Front Pediatr, 2016. **4**: p. 121.
8. Trevarthen, C. and J.T. Delafield-Butt, *Autism as a developmental disorder in intentional movement and affective engagement.* Front Integr Neurosci, 2013. **7**: p. 49.
9. Prechtl, H.F.R., *Continuity of neural functions from prenatal to postnatal life*. Clinics in developmental medicine. 1984, London, Philadelphia: Spastics International Medical Publications, J.B. Lippincott Co. viii, 255 p.
10. Mohamed Thangal, S.N. and J.M. Donelan, *Scaling of inertial delays in terrestrial mammals.* PLoS One, 2020. **15**(2): p. e0217188.
11. Moore, J.K., L.M. Perazzo, and A. Braun, *Time course of axonal myelination in the human brainstem auditory pathway.* Hear Res, 1995. **87**(1-2): p. 21-31.
12. More, H.L. and J.M. Donelan, *Scaling of sensorimotor delays in terrestrial mammals.* Proc Biol Sci, 2018. **285**(1885).
13. More, H.L., et al., *Scaling of sensorimotor control in terrestrial mammals.* Proc Biol Sci, 2010. **277**(1700): p. 3563-8.
14. Brown, D.K. and J.C. Dort, *Auditory neuropathy: when test results conflict.* J Otolaryngol, 2001. **30**(1): p. 46-51.
15. Starr, A., et al., *Development of auditory function in newborn infants revealed by auditory brainstem potentials.* Pediatrics, 1977. **60**(6): p. 831-9.
16. Stiles, J. and T.L. Jernigan, *The basics of brain development.* Neuropsychol Rev, 2010. **20**(4): p. 327-48.
17. Torres, E.B., *Two classes of movements in motor control.* Exp Brain Res, 2011. **215**(3-4): p. 269-83.
18. Torres, E.B., *Signatures of movement variability anticipate hand speed according to levels of intent.* Behav Brain Funct, 2013. **9**: p. 10.
19. Torres, E.B., P. Yanovich, and D.N. Metaxas, *Give spontaneity and self-discovery a chance in ASD: spontaneous peripheral limb variability as a proxy to evoke centrally driven intentional acts.* Front Integr Neurosci, 2013. **7**: p. 46.
20. Baizer, J.S., *Functional and Neuropathological Evidence for a Role of the Brainstem in Autism.* Front Integr Neurosci, 2021. **15**: p. 748977.

Not peer review

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

21. Jacob, S.N. and H. Nienborg, *Monoaminergic Neuromodulation of Sensory Processing.* Front Neural Circuits, 2018. **12**: p. 51.
22. Kandler, K., A. Clause, and J. Noh, *Tonotopic reorganization of developing auditory brainstem circuits.* Nat Neurosci, 2009. **12**(6): p. 711-7.
23. Clause, A., et al., *The precise temporal pattern of prehearing spontaneous activity is necessary for tonotopic map refinement.* Neuron, 2014. **82**(4): p. 822-35.
24. Friauf, E. and C. Lohmann, *Development of auditory brainstem circuitry. Activity-dependent and activity-independent processes.* Cell Tissue Res, 1999. **297**(2): p. 187-95.
25. Inui, T., S. Kumagaya, and M. Myowa-Yamakoshi, *Neurodevelopmental Hypothesis about the Etiology of Autism Spectrum Disorders.* Front Hum Neurosci, 2017. **11**: p. 354.
26. Dwornicka, B., et al., *Attempt of Determining the Fetal Reaction to Acoustic Stimulation.* Acta Otolaryngol, 1964. **57**: p. 571-4.
27. Slaninova, B., K. Guth, and K. Mil, *[Determination of the reaction of the fetus to acoustic stimulation].* Cas Lek Cesk, 1968. **107**(36): p. 1101-3.
28. Sano, M., et al., *Early myelination patterns in the brainstem auditory nuclei and pathway: MRI evaluation study.* Int J Pediatr Otorhinolaryngol, 2007. **71**(7): p. 1105-15.
29. Miron, O., et al., *Prolonged Auditory Brainstem Response in Universal Hearing Screening of Newborns with Autism Spectrum Disorder.* Autism Res, 2021. **14**(1): p. 46-52.
30. Seethapathy, J., et al., *Auditory brainstem response in very preterm, moderately preterm and late preterm infants.* Int J Pediatr Otorhinolaryngol, 2018. **111**: p. 119-127.
31. Torres, E.B., *Precision Autism: Genomic Stratification of Disorders Making Up the Broad Spectrum May Demystify its "Epidemic Rates".* bioRxvi, 2021: p. 28.
32. Consortium, G.T., *The GTEx Consortium atlas of genetic regulatory effects across human tissues.* Science, 2020. **369**(6509): p. 1318-1330.
33. Wolpert, D.M., R.C. Miall, and M. Kawato, *Internal models in the cerebellum.* Trends Cogn Sci, 1998. **2**(9): p. 338-47.
34. Yao, Z., et al., *A Single-Cell Roadmap of Lineage Bifurcation in Human ESC Models of Embryonic Brain Development.* Cell Stem Cell, 2017. **20**(1): p. 120-134.
35. Chow, C.K. and C.H. Liu, *Approximating discrete probability distributions with dependence trees.* IEEE Trans Inform Theory IT, 1968. **14**: p. 462-67.
36. Inoue, K., Y. Tanabe, and J.R. Lupski, *Myelin deficiencies in both the central and the peripheral nervous systems associated with a SOX10 mutation.* Ann Neurol, 1999. **46**(3): p. 313-8.
37. Balter, M., *Language evolution. 'Speech gene' tied to modern humans.* Science, 2002. **297**(5584): p. 1105.
38. Belton, E., et al., *Bilateral brain abnormalities associated with dominantly inherited verbal and orofacial dyspraxia.* Hum Brain Mapp, 2003. **18**(3): p. 194-200.
39. Enard, W., et al., *Molecular evolution of FOXP2, a gene involved in speech and language.* Nature, 2002. **418**(6900): p. 869-72.
40. Gauthier, J., et al., *Truncating mutations in NRXN2 and NRXN1 in autism spectrum disorders and schizophrenia.* Hum Genet, 2011. **130**(4): p. 563-73.
41. Reichelt, A.C., R.J. Rodgers, and S.J. Clapcote, *The role of neurexins in schizophrenia and autistic spectrum disorder.* Neuropharmacology, 2012. **62**(3): p. 1519-26.

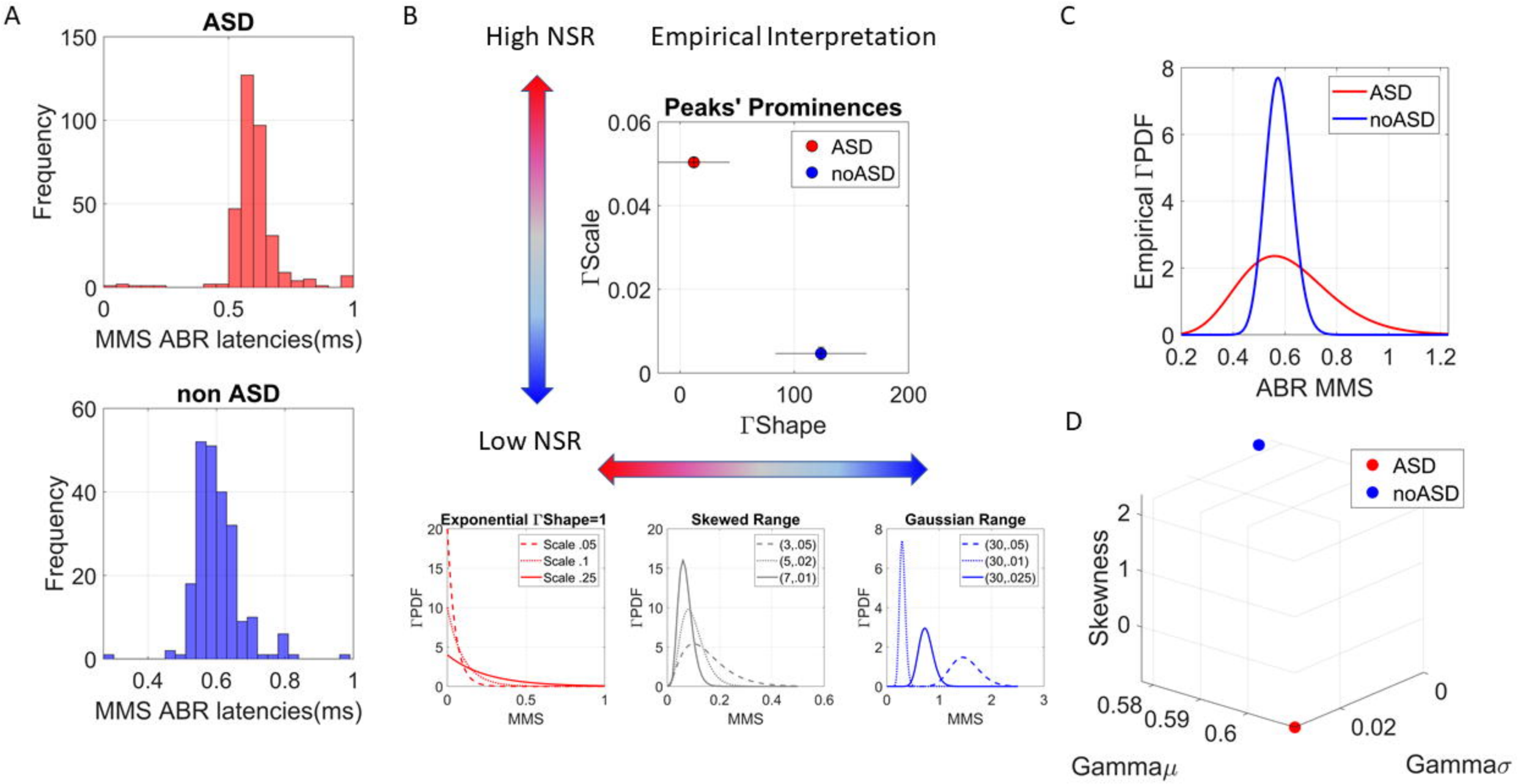
bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

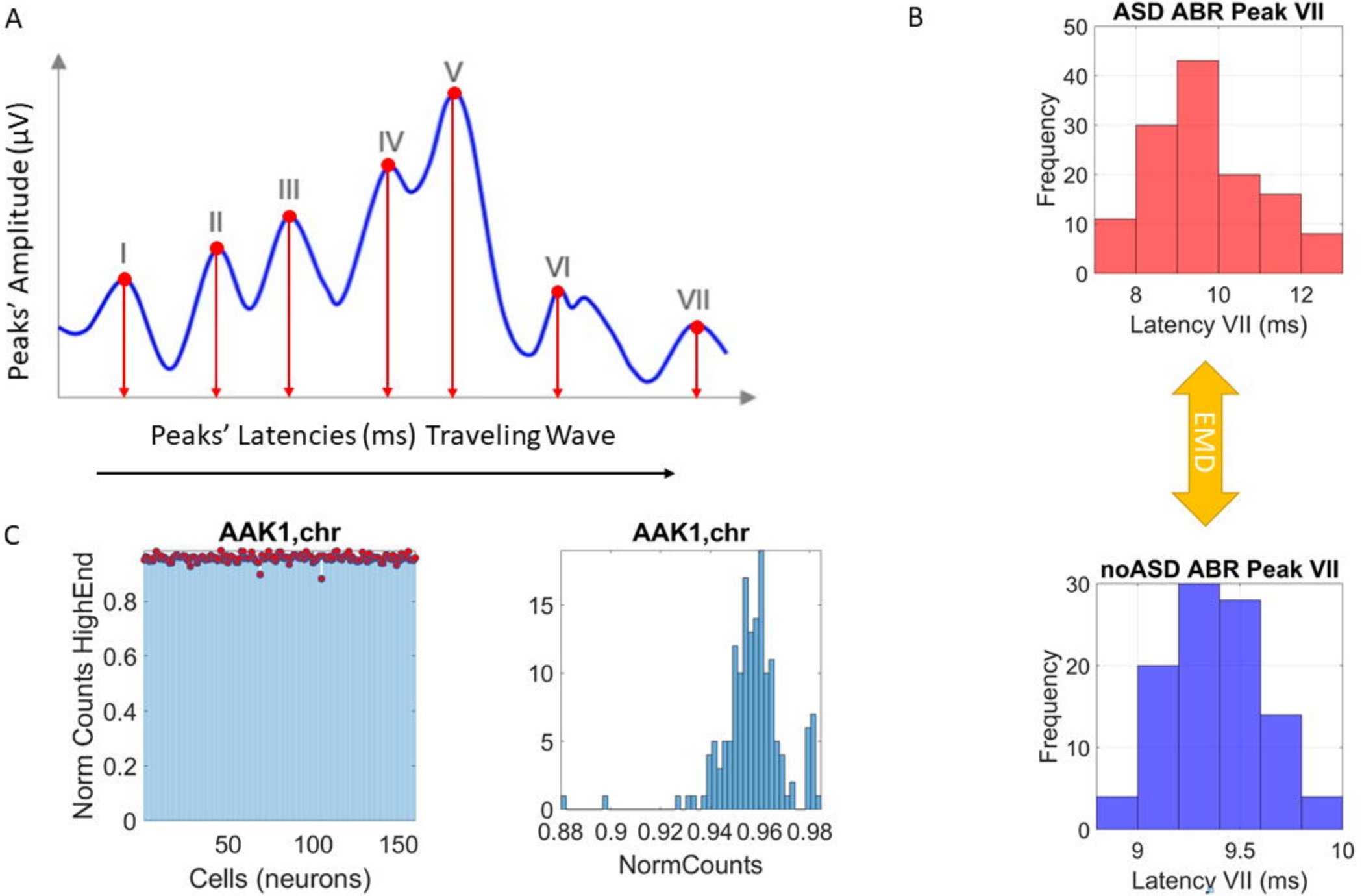
42. Rujescu, D., et al., *Disruption of the neurexin 1 gene is associated with schizophrenia.* Hum Mol Genet, 2009. **18**(5): p. 988-96.
43. Taylor, S.C., et al., *The Role of Synaptic Cell Adhesion Molecules and Associated Scaffolding Proteins in Social Affiliative Behaviors.* Biol Psychiatry, 2020. **88**(6): p. 442-451.
44. Rowen, L., et al., *Analysis of the human neurexin genes: alternative splicing and the generation of protein diversity.* Genomics, 2002. **79**(4): p. 587-97.
45. Graf, E.R., et al., *Neurexins induce differentiation of GABA and glutamate postsynaptic specializations via neuroligins.* Cell, 2004. **119**(7): p. 1013-26.
46. Reissner, C., et al., *Mutational analysis of the neurexin/neuroligin complex reveals essential and regulatory components.* Proc Natl Acad Sci U S A, 2008. **105**(39): p. 15124-9.
47. Karmel, B.Z., et al., *Brain-stem auditory evoked responses as indicators of early brain insult.* Electroencephalogr Clin Neurophysiol, 1988. **71**(6): p. 429-42.
48. Seif, A., et al., *A Systematic Review of Brainstem Contributions to Autism Spectrum Disorder.* Front Integr Neurosci, 2021. **15**: p. 760116.
49. Hawgood, S., et al., *Precision medicine: Beyond the inflection point.* Sci Transl Med, 2015. **7**(300): p. 300ps17.
50. Vaskevich, A. and E.B. Torres, *Rethinking Statistical Learning as a Dynamic Stochastic Process, from The Motor Systems Perspective.* bioRxvi, 2022.
51. Ponton, C.W., et al., *Frequency-specific maturation of the eighth nerve and brain-stem auditory pathway: evidence from derived auditory brain-stem responses (ABRs).* J Acoust Soc Am, 1992. **91**(3): p. 1576-86.
52. Beintema, D.J., *A neurological study of newborn infants*. Clinics in developmental medicine,. 1968, London,: Spastics International Medical Publications; London, Distributed by Heinemann Medical. 11 , 178 p.
53. Touwen, B.C.L. and H.F.R. Prechtl, *The neurological examination of the child with minor nervous dysfunction*. Clinics in developmental medicine, no 38. 1970, London, Philadelphia,: Heinemann Medical, Lippincott for Spastics International Medical Publications. 7 , vi, 105 p.
54. Prechtl, H.F.R., *The neurological examination of the full term newborn infant*. 2nd ed. Clinics in developmental medicine no 63. 1977, London, Philadelphia: Heinemann Medical for Spastics International Medical Publications, Lippincott. viii, 68 p.
55. Luyster, R., et al., *The Autism Diagnostic Observation Schedule-toddler module: a new module of a standardized diagnostic measure for autism spectrum disorders.* J Autism Dev Disord, 2009. **39**(9): p. 1305-20.
56. Lord, C., et al., *Autism diagnostic observation schedule: a standardized observation of communicative and social behavior.* J Autism Dev Disord, 1989. **19**(2): p. 185-212.
57. Lord, C., et al., *The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism.* J Autism Dev Disord, 2000. **30**(3): p. 205-23.
58. Torres, E.B. and C. Whyatt, *Autism : the movement sensing perspective*. 2018, Boca Raton: CRC Press/Taylor & Francis Group. xviii, 386 pages.
59. Bokadia, H., R. Rai, and E.B. Torres, *Digitized Autism Observation Diagnostic Schedule: Social Interactions beyond the Limits of the Naked Eye.* J Pers Med, 2020. **10**(4).

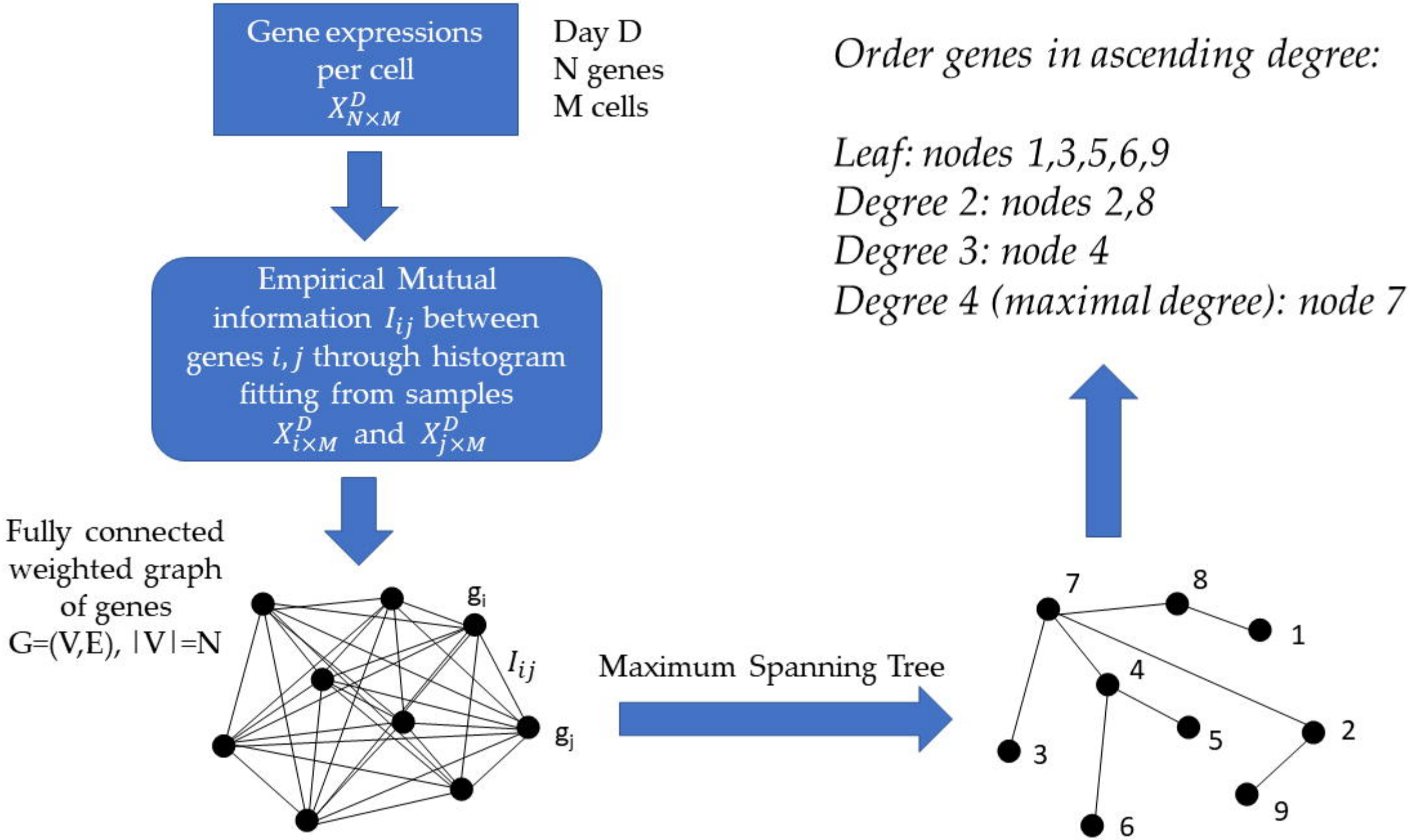
bioRxiv preprint doi: https://doi.org/10.1101/2022.01.27.478048; this version posted January 28, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

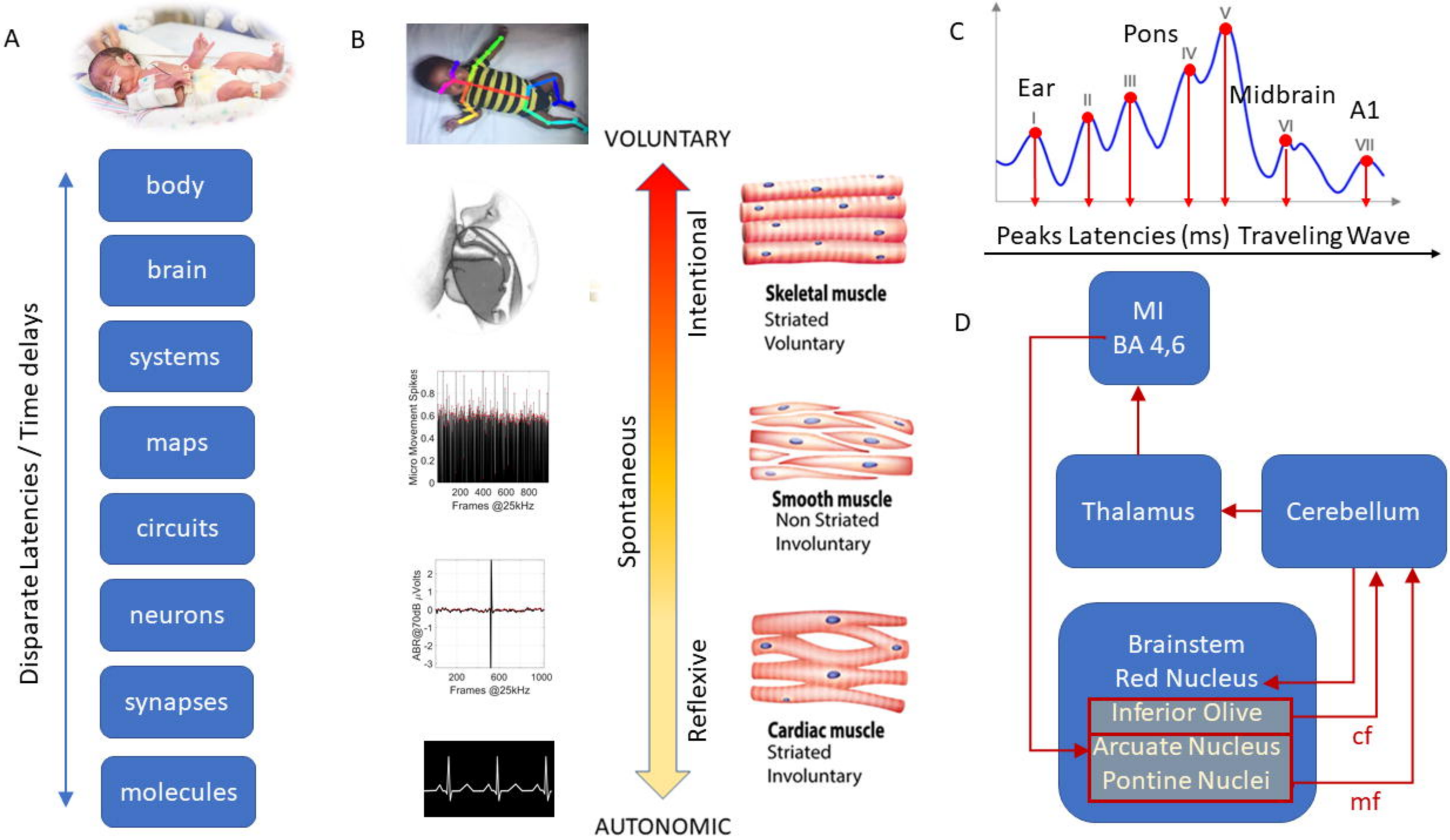
60. Torres, E.B., et al., *Hidden Aspects of the Research ADOS Are Bound to Affect Autism Science.* Neural Comput, 2020. **32**(3): p. 515-561.
61. Kittler, P.M., et al., *The development of selective attention and inhibition in NICU graduates during the preschool years.* Dev Neuropsychol, 2011. **36**(8): p. 1003-17.
62. Karmel, B.Z., et al., *Early medical and behavioral characteristics of NICU infants later classified with ASD.* Pediatrics, 2010. **126**(3): p. 457-67.
63. Cohen, I.L. and V. Sudhalter, *The PDD Behavior Inventory*, in *Lutz, Fl*. 2005, Psychological Assessment Resources, Inc.: New York.
64. Torres, E.B., *Methods for the diagnosis and treatment of neurological disorders*, in *Google Patents*, USPTO, Editor. 2018, Rutgers State University of New Jersey: US. p. 38.
65. Monge, G., *Memoire sur la theorie des deblais et des remblais.*, in *Histoire de l' Academie Royale des Science; avec les Memoires de Mathematique et de Physique;* . 1781, De L'imprimerie Royale: Paris, France.
66. Rubner, Y., C. Tomasi, and L.J. Guibas. *Metric for Distributions with Applications to Image Databases.* in *Proceedings of the ICCV*. 1998. Bombay, India.
67. Gretton, A., et al. *A Kernel Statistical Test of Independence*. in *Advances in Neural Information Processing Systems 20 (NeurIPS 2007)*. 2007. Vancouver, B.C., Canada.

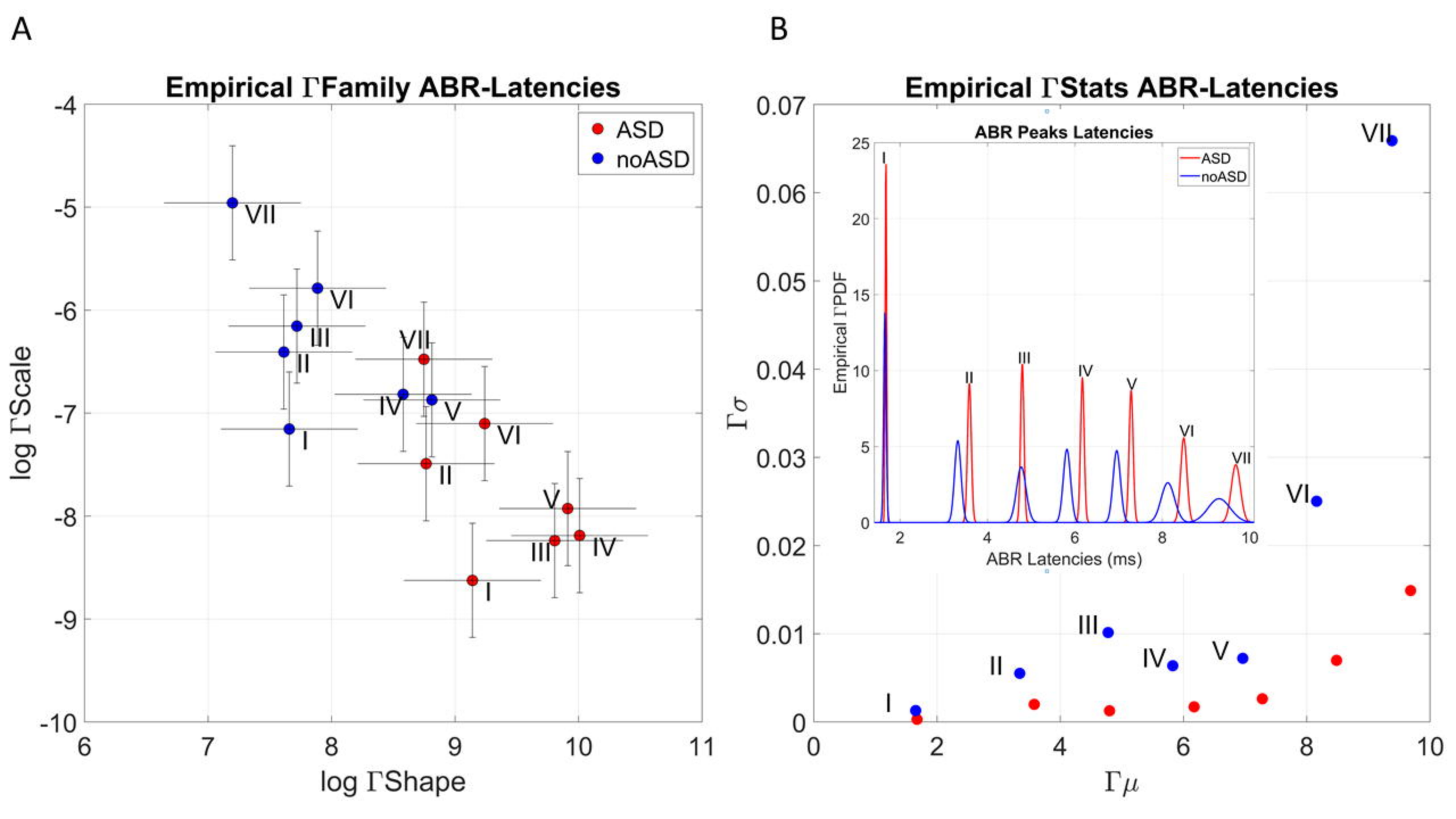
For Peer Review

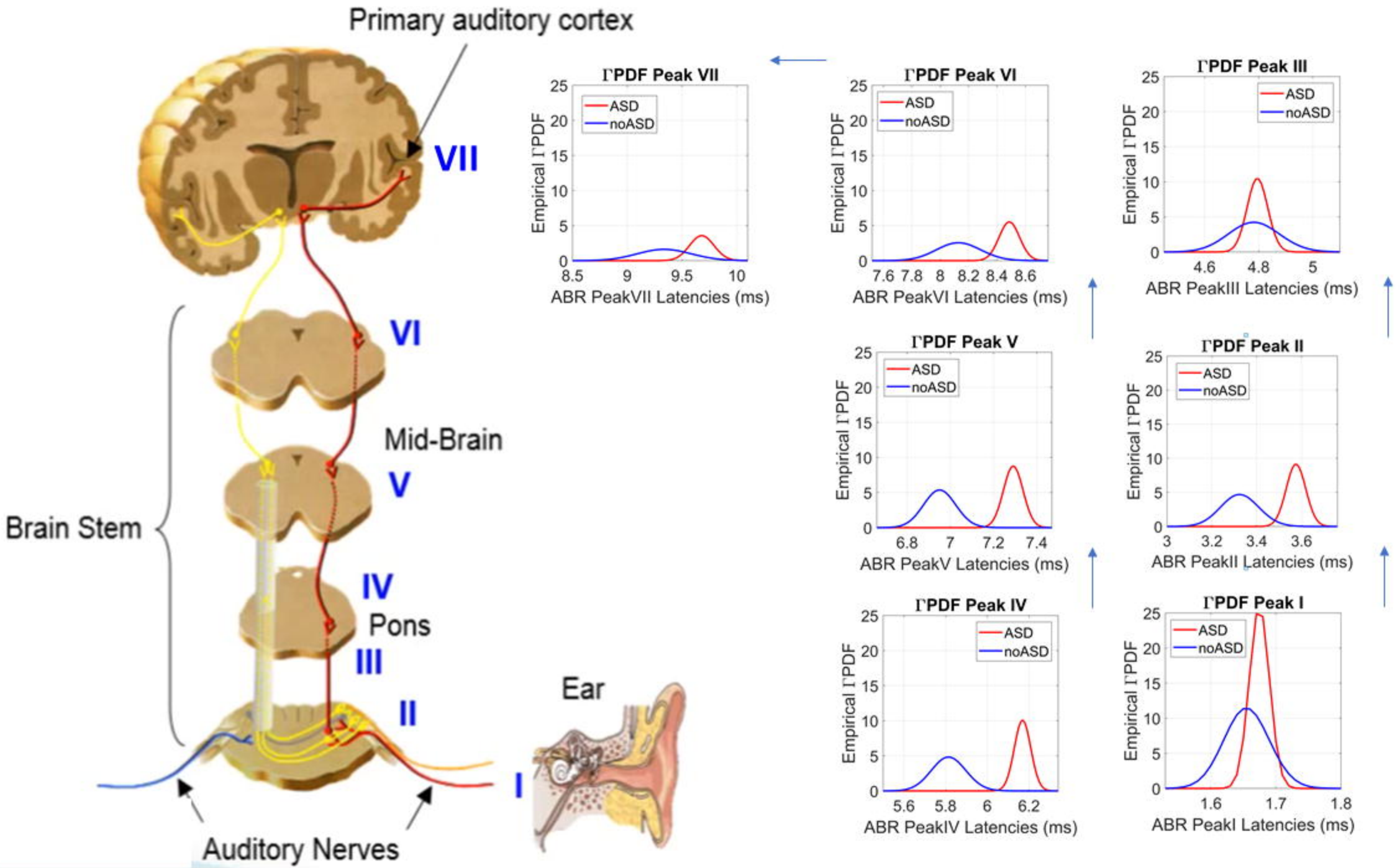

A
ASD
Frequency
MMS ABR latencies(ms)
non ASD
Frequency
MMS ABR latencies(ms)
B
High NSR
Empirical Interpretation
Peaks' Prominences
ASD
noASD
ΓScale
ΓShape
Low NSR
Exponential ΓShape=1
Scale .05
Scale .1
Scale .25
ΓPDF
MMS
Skewed Range
(3,.05)
(5,.02)
(7,.01)
Gaussian Range
(30,.05)
(30,.01)
(30,.025)
C
Empirical ΓPDF
ASD
noASD
ABR MMS
D
ASD
noASD
Skewness
Gammaμ
Gammaσ

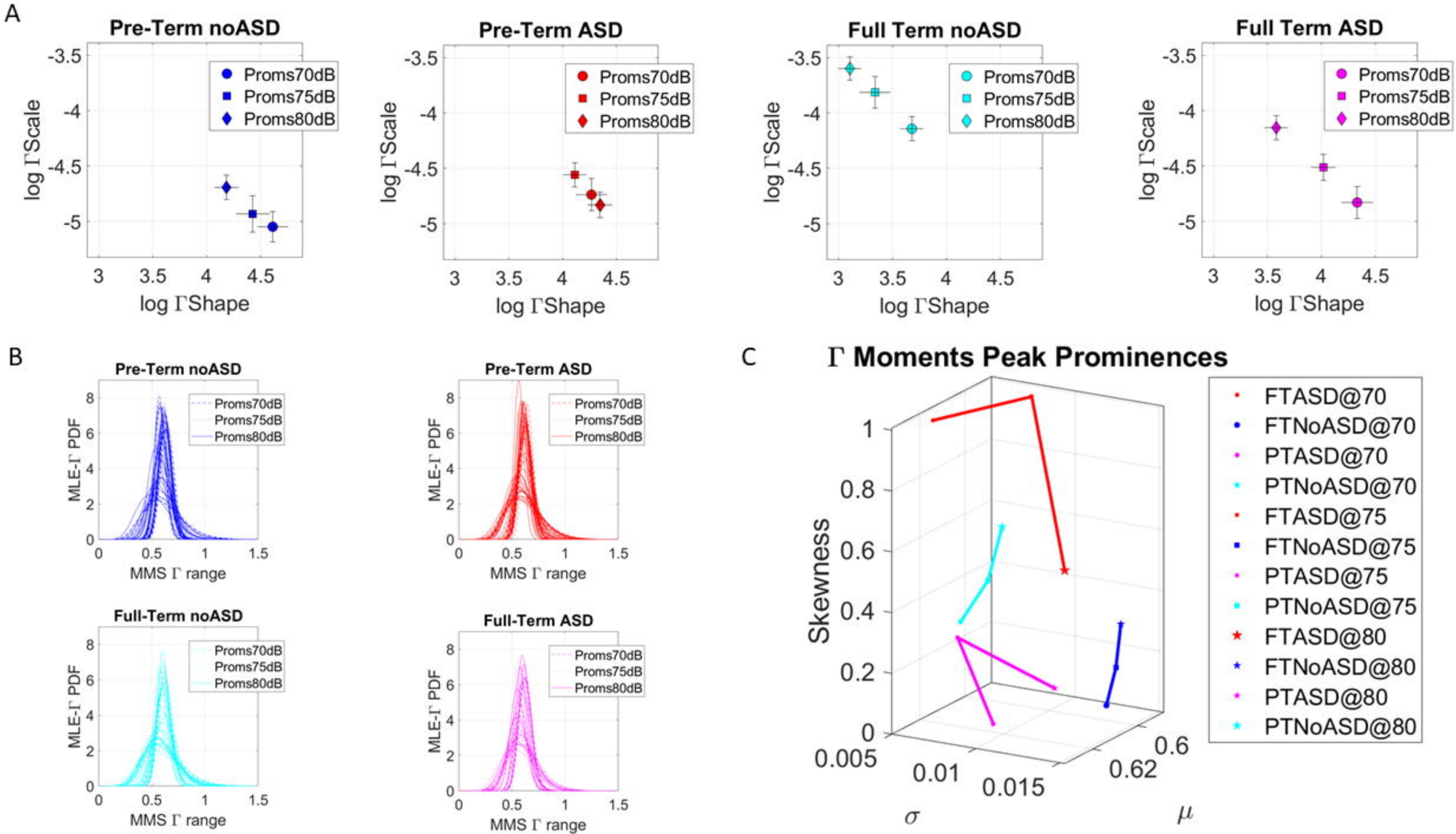

A
Peaks' Amplitude (µV)
I
II
III
IV
V
VI
VII
Peaks' Latencies (ms) Traveling Wave
B
ASD ABR Peak VII
Frequency
Latency VII (ms)
EMD
noASD ABR Peak VII
Frequency
Latency VII (ms)
C
AAK1,chr
Norm Counts HighEnd
Cells (neurons)
AAK1,chr
NormCounts

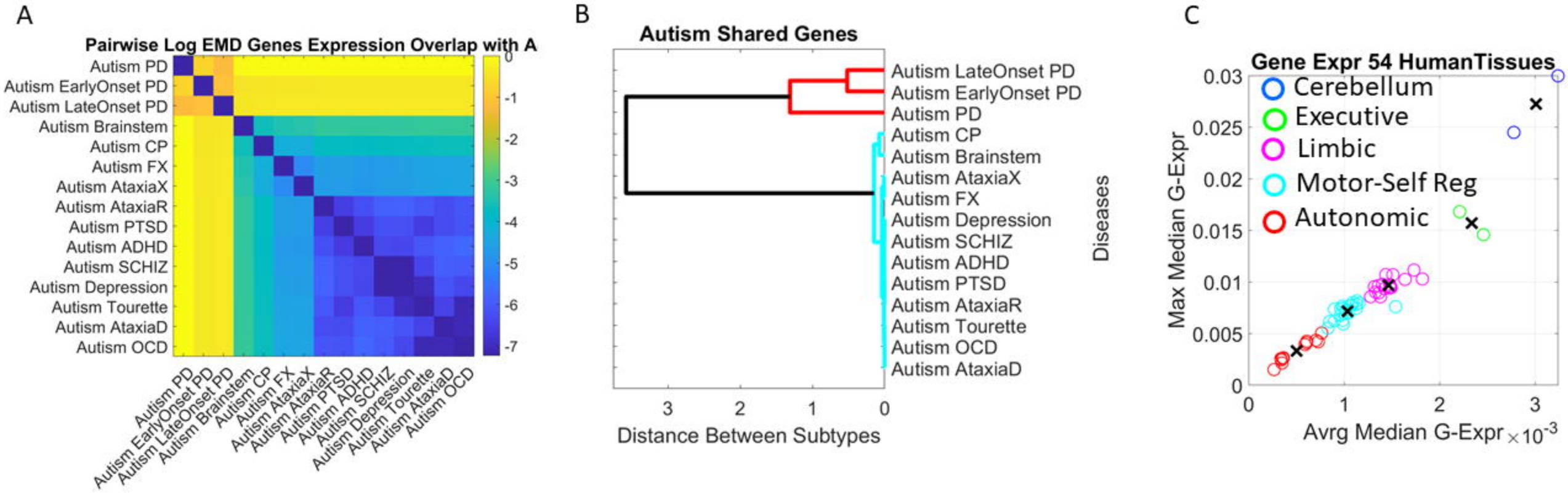

Gene expressions per cell $X^D_{N\times M}$
Day D
N genes
M cells
Empirical Mutual information I_{ij} between genes i, j through histogram fitting from samples $X^D_{i\times M}$ and $X^D_{j\times M}$
Fully connected weighted graph of genes G=(V,E), |V|=N
g_i
I_{ij}
g_j
Maximum Spanning Tree
7
8
1
4
5
2
3
9
6
Order genes in ascending degree:
Leaf: nodes 1,3,5,6,9
Degree 2: nodes 2,8
Degree 3: node 4
Degree 4 (maximal degree): node 7

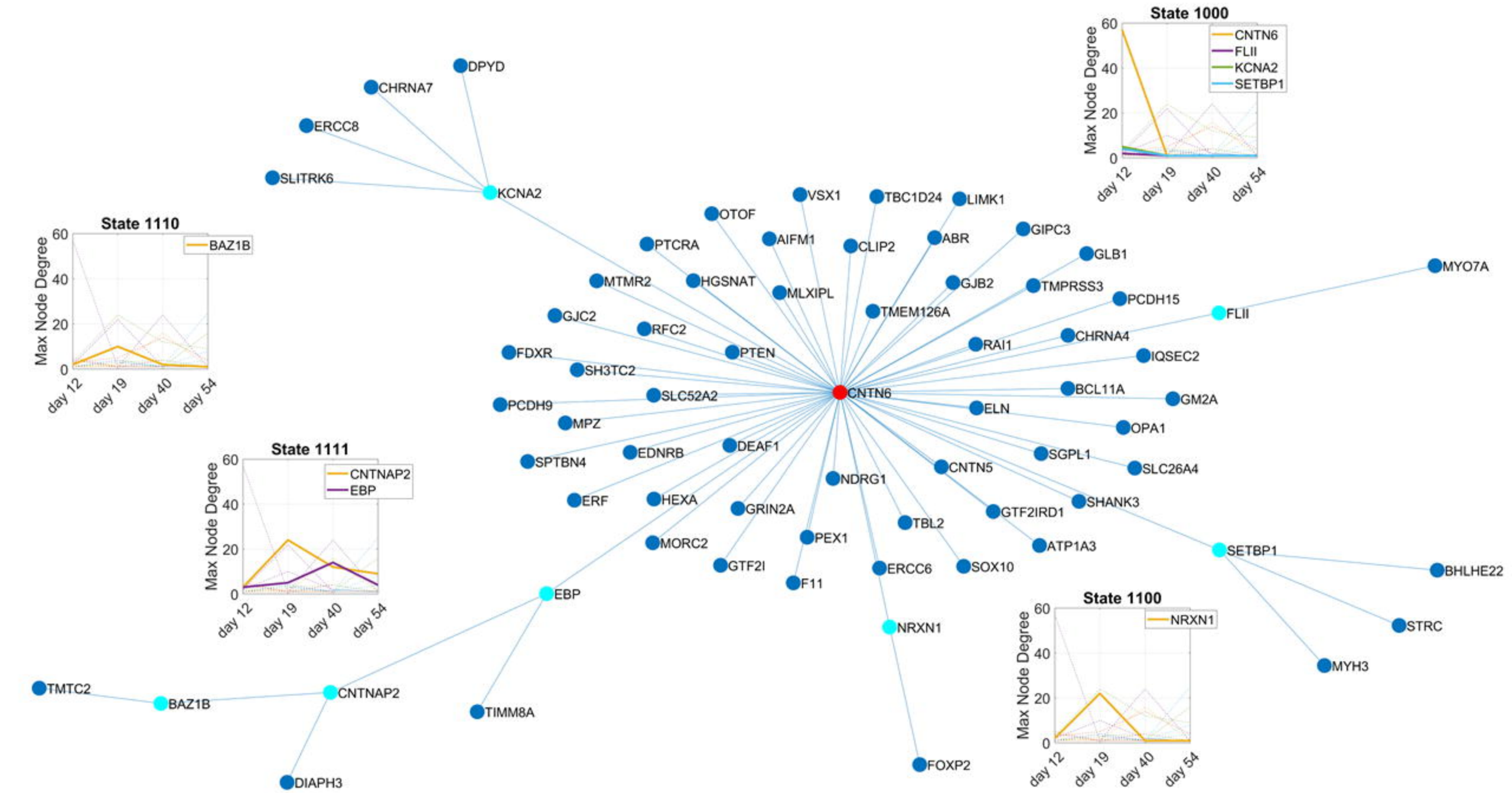

A
Disparate Latencies / Time delays
body
brain
systems
maps
circuits
neurons
synapses
molecules
B
Micro Movement Spikes
Frames @25kHz
ABR@70dB µVolts
Frames @25kHz
VOLUNTARY
Intentional
Spontaneous
Reflexive
AUTONOMIC
Skeletal muscle
Striated
Voluntary
Smooth muscle
Non Striated
Involuntary
Cardiac muscle
Striated
Involuntary
C
Ear
Pons
Midbrain
A1
I
II
III
IV
V
VI
VII
Peaks Latencies (ms) Traveling Wave
D
MI
BA 4,6
Thalamus
Cerebellum
Brainstem
Red Nucleus
Inferior Olive
Arcuate Nucleus
Pontine Nuclei
cf
mf

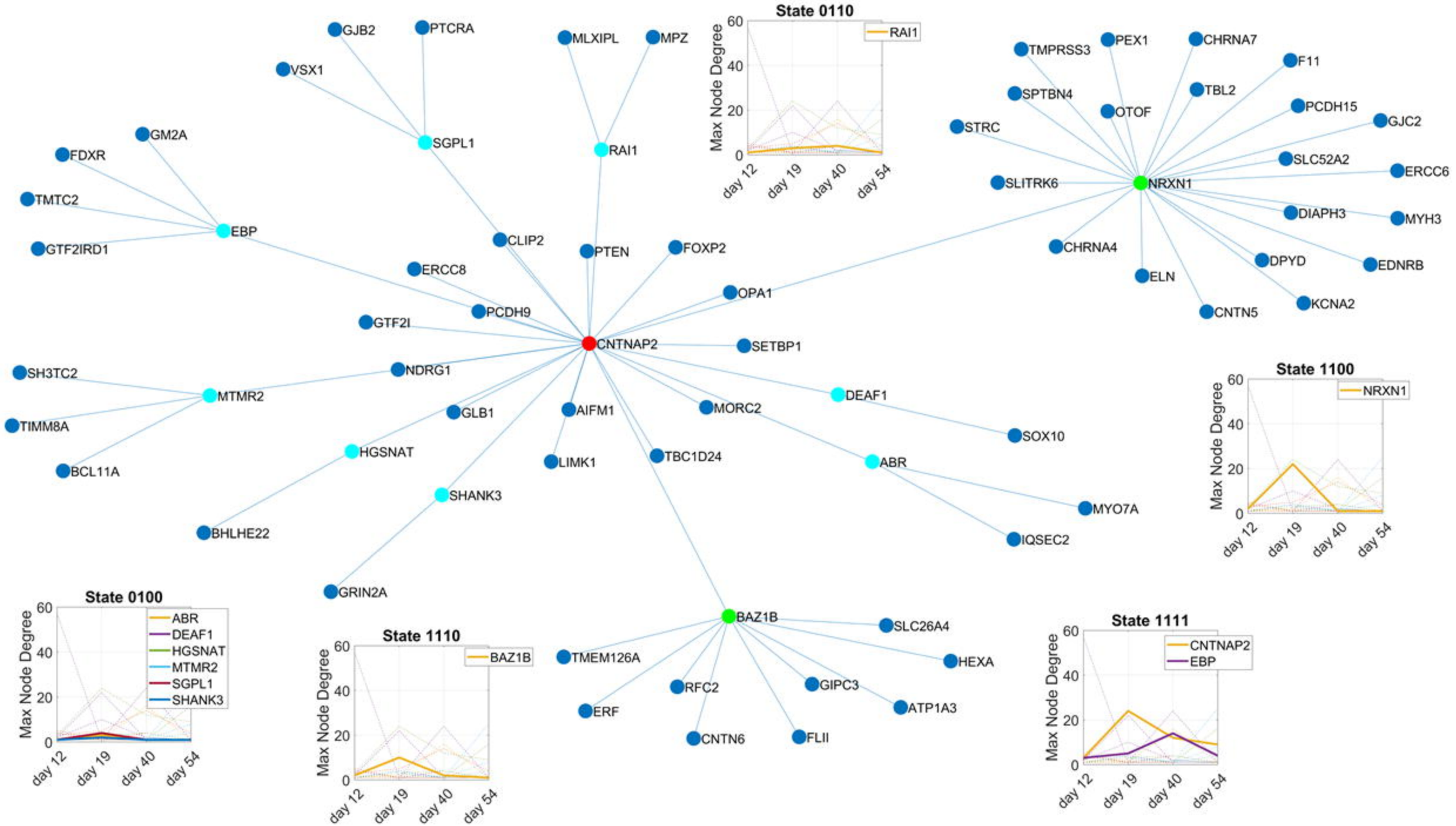

A
Empirical ΓFamily ABR-Latencies
ASD
noASD
log ΓScale
log ΓShape
B
Empirical ΓStats ABR-Latencies
ABR Peaks Latencies
ASD
noASD
Empirical ΓPDF
ABR Latencies (ms)
Γσ
Γμ

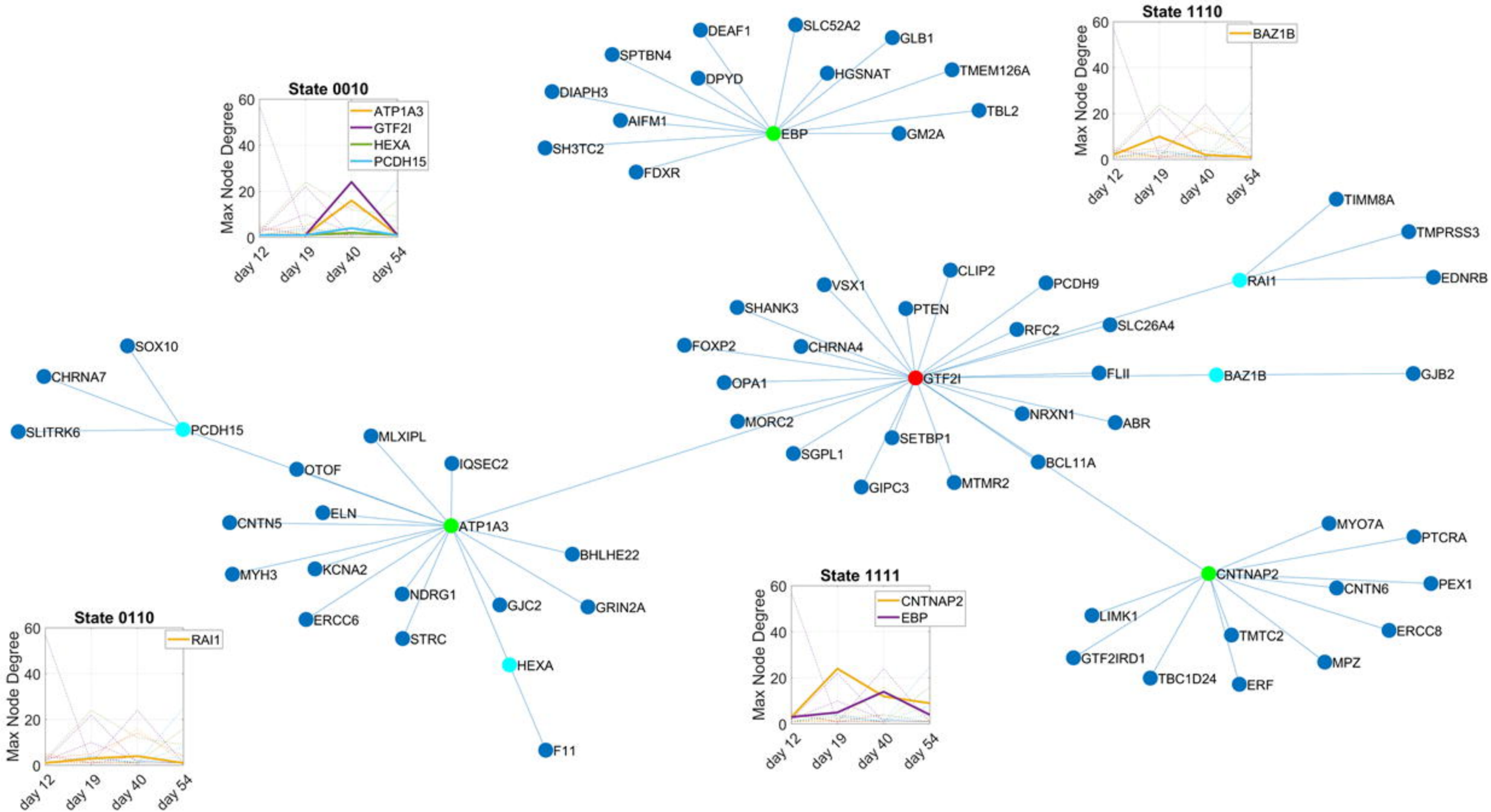

Primary auditory cortex
VII
VI
Mid-Brain
V
Brain Stem
IV
Pons
III
II
I
Auditory Nerves
Ear
ΓPDF Peak VII
ΓPDF Peak VI
ΓPDF Peak III
ΓPDF Peak V
ΓPDF Peak II
ΓPDF Peak IV
ΓPDF Peak I
ASD
noASD
Empirical ΓPDF
ABR PeakVII Latencies (ms)
ABR PeakVI Latencies (ms)
ABR PeakIII Latencies (ms)
ABR PeakV Latencies (ms)
ABR PeakII Latencies (ms)
ABR PeakIV Latencies (ms)
ABR PeakI Latencies (ms)

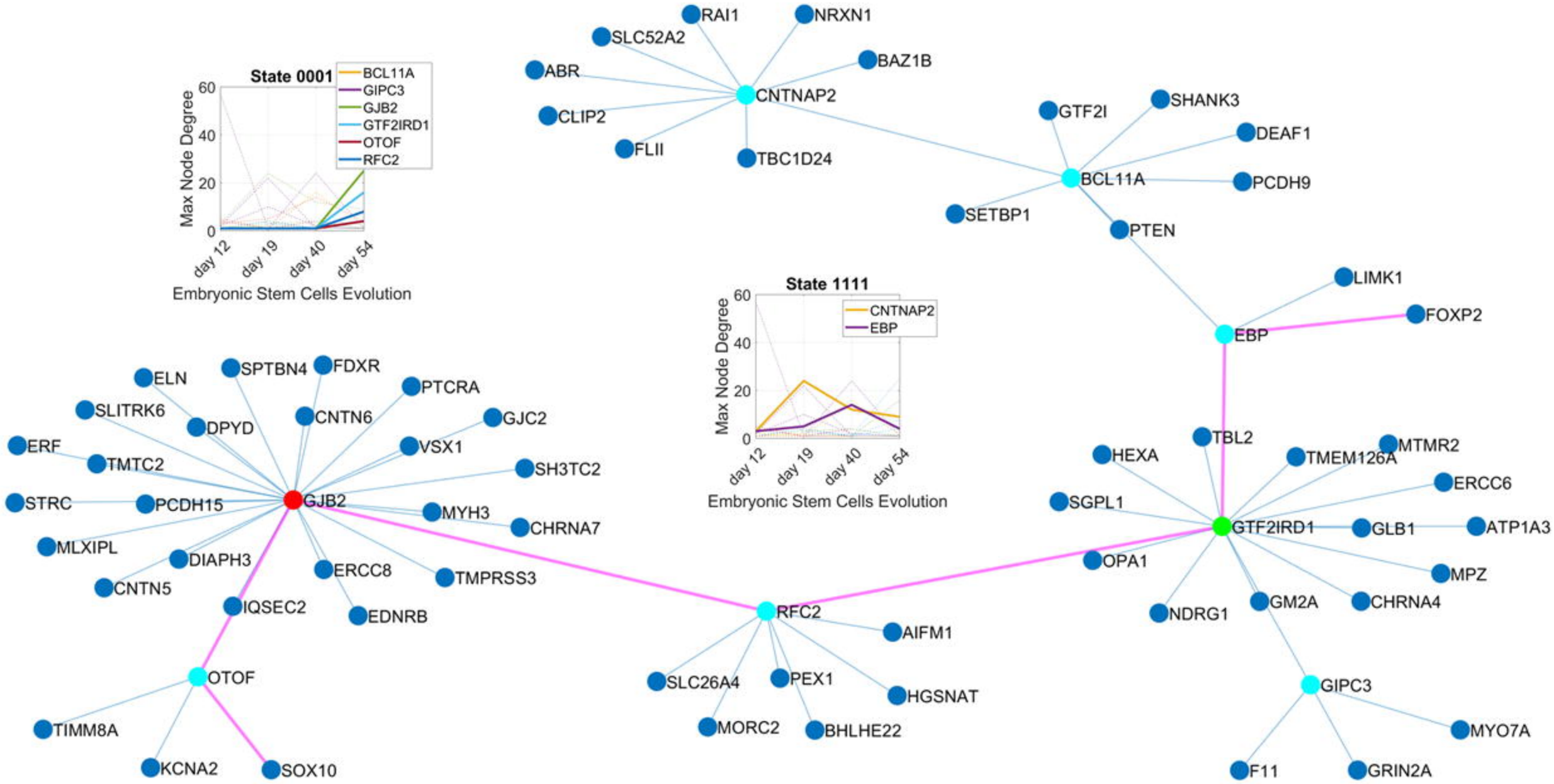

A
Pre-Term noASD
Pre-Term ASD
Full Term noASD
Full Term ASD
Proms70dB
Proms75dB
Proms80dB
log ΓScale
log ΓShape
B
Pre-Term noASD
Pre-Term ASD
Full-Term noASD
Full-Term ASD
MLE-Γ PDF
MMS Γ range
C
Γ Moments Peak Prominences
Skewness
σ
μ
FTASD@70
FTNoASD@70
PTASD@70
PTNoASD@70
FTASD@75
FTNoASD@75
PTASD@75
PTNoASD@75
FTASD@80
FTNoASD@80
PTASD@80
PTNoASD@80

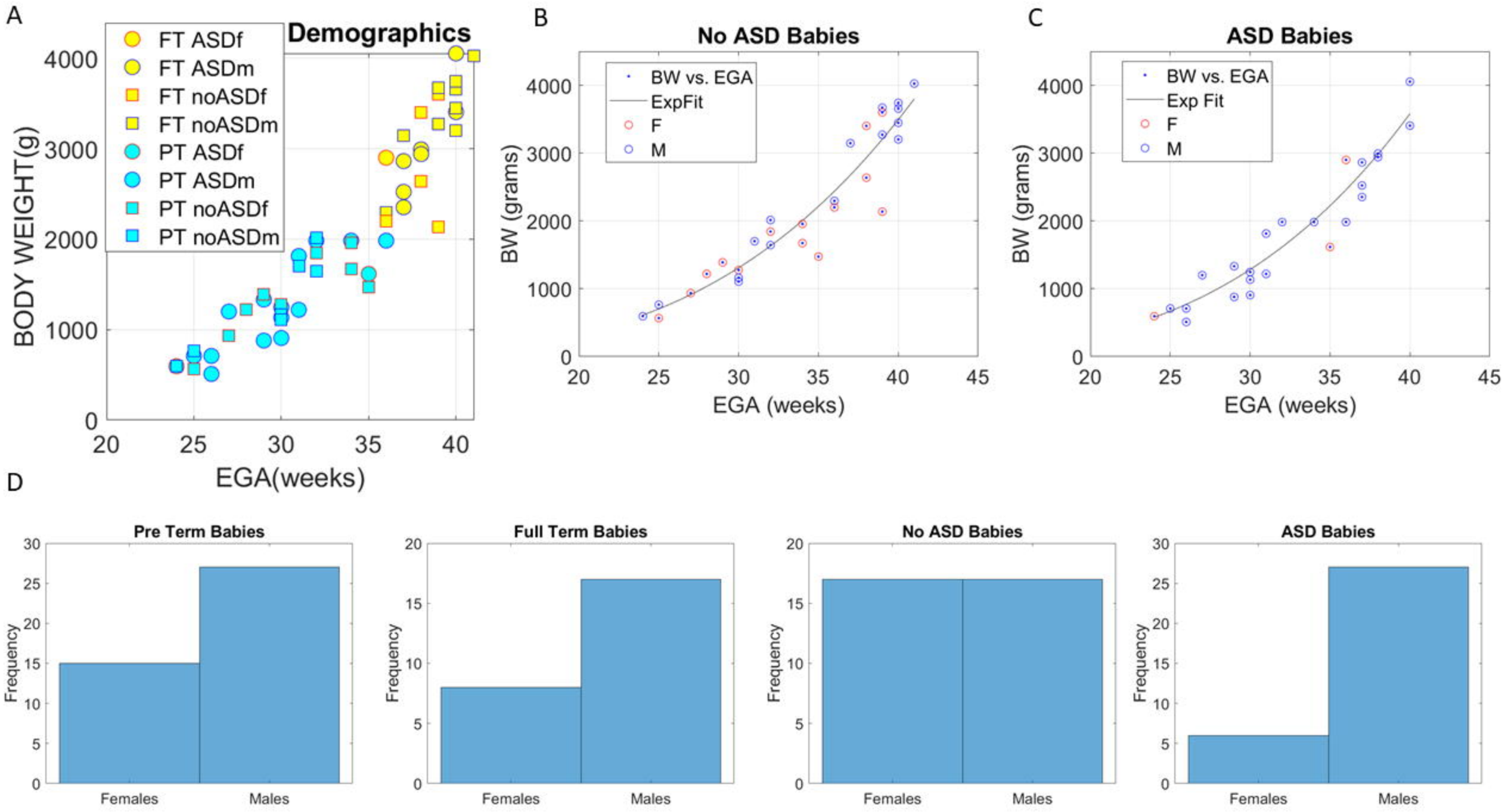

A
Pairwise Log EMD Genes Expression Overlap with A
Autism PD
Autism EarlyOnset PD
Autism LateOnset PD
Autism Brainstem
Autism CP
Autism FX
Autism AtaxiaX
Autism AtaxiaR
Autism PTSD
Autism ADHD
Autism SCHIZ
Autism Depression
Autism Tourette
Autism AtaxiaD
Autism OCD
0
-1
-2
-3
-4
-5
-6
-7
B
Autism Shared Genes
Autism LateOnset PD
Autism EarlyOnset PD
Autism PD
Autism CP
Autism Brainstem
Autism AtaxiaX
Autism FX
Autism Depression
Autism SCHIZ
Autism ADHD
Autism PTSD
Autism AtaxiaR
Autism Tourette
Autism OCD
Autism AtaxiaD
Diseases
3
2
1
0
Distance Between Subtypes
C
Gene Expr 54 HumanTissues
Cerebellum
Executive
Limbic
Motor-Self Reg
Autonomic
Max Median G-Expr
0.03
0.025
0.02
0.015
0.01
0.005
0
0
1
2
3
Avrg Median G-Expr ×10^-3

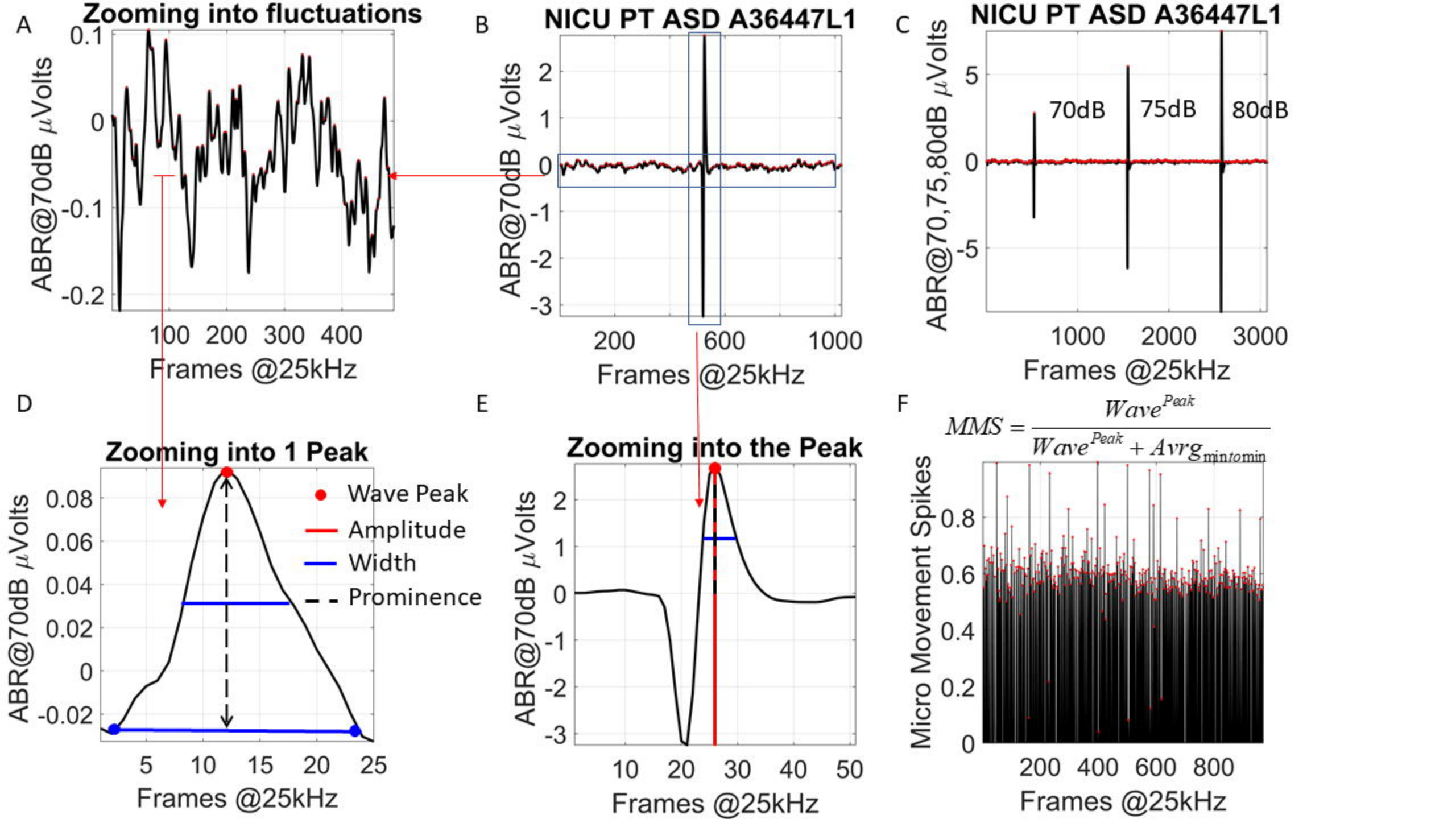

State 1000
Max Node Degree
CNTN6
FLII
KCNA2
SETBP1
day 12
day 19
day 40
day 54
State 1110
BAZ1B
State 1111
CNTNAP2
EBP
State 1100
NRXN1
DPYD
CHRNA7
ERCC8
SLITRK6
KCNA2
VSX1
TBC1D24
LIMK1
OTOF
AIFM1
CLIP2
ABR
GIPC3
PTCRA
GLB1
MTMR2
HGSNAT
MLXIPL
GJB2
TMPRSS3
MYO7A
PCDH15
GJC2
TMEM126A
FLII
RFC2
RAI1
CHRNA4
FDXR
PTEN
IQSEC2
SH3TC2
CNTN6
BCL11A
SLC52A2
GM2A
PCDH9
ELN
MPZ
OPA1
EDNRB
DEAF1
SGPL1
SPTBN4
CNTN5
SLC26A4
NDRG1
ERF
HEXA
SHANK3
GRIN2A
GTF2IRD1
TBL2
MORC2
PEX1
ATP1A3
SETBP1
GTF2I
ERCC6
SOX10
BHLHE22
F11
EBP
NRXN1
STRC
MYH3
TMTC2
BAZ1B
CNTNAP2
TIMM8A
FOXP2
DIAPH3


GJB2
PTCRA
MLXIPL
MPZ
VSX1
SGPL1
RAI1
GM2A
FDXR
TMTC2
EBP
GTF2IRD1
CLIP2
PTEN
FOXP2
ERCC8
PCDH9
GTF2I
OPA1
CNTNAP2
SETBP1
SH3TC2
NDRG1
MTMR2
TIMM8A
GLB1
AIFM1
MORC2
DEAF1
SOX10
HGSNAT
LIMK1
TBC1D24
ABR
BCL11A
SHANK3
MYO7A
BHLHE22
IQSEC2
GRIN2A
BAZ1B
SLC26A4
TMEM126A
HEXA
RFC2
GIPC3
ERF
ATP1A3
CNTN6
FLII
TMPRSS3
PEX1
CHRNA7
F11
SPTBN4
TBL2
OTOF
PCDH15
STRC
GJC2
SLC52A2
SLITRK6
NRXN1
ERCC6
DIAPH3
MYH3
CHRNA4
DPYD
EDNRB
ELN
CNTN5
KCNA2
State 0110
State 1100
State 0100
State 1110
State 1111
Max Node Degree
0
20
40
60
day 12
day 19
day 40
day 54
RAI1
NRXN1
ABR
DEAF1
HGSNAT
MTMR2
SGPL1
SHANK3
BAZ1B
CNTNAP2
EBP


State 0010
Max Node Degree
ATP1A3
GTF2I
HEXA
PCDH15
day 12
day 19
day 40
day 54
State 1110
BAZ1B
State 0110
RAI1
State 1111
CNTNAP2
EBP
EBP
DEAF1
SLC52A2
GLB1
SPTBN4
DPYD
HGSNAT
TMEM126A
DIAPH3
TBL2
AIFM1
GM2A
SH3TC2
FDXR
GTF2I
CLIP2
VSX1
PCDH9
SHANK3
PTEN
RFC2
SLC26A4
FOXP2
CHRNA4
FLII
BAZ1B
GJB2
OPA1
NRXN1
ABR
MORC2
SETBP1
SGPL1
BCL11A
GIPC3
MTMR2
RAI1
TIMM8A
TMPRSS3
EDNRB
PCDH15
SOX10
CHRNA7
SLITRK6
ATP1A3
MLXIPL
IQSEC2
OTOF
ELN
CNTN5
BHLHE22
MYH3
KCNA2
NDRG1
GJC2
GRIN2A
ERCC6
STRC
HEXA
F11
CNTNAP2
MYO7A
PTCRA
CNTN6
PEX1
LIMK1
ERCC8
TMTC2
GTF2IRD1
MPZ
TBC1D24
ERF


State 0001
BCL11A
GIPC3
GJB2
GTF2IRD1
OTOF
RFC2
Max Node Degree
60
40
20
0
day 12
day 19
day 40
day 54
Embryonic Stem Cells Evolution
State 1111
CNTNAP2
EBP
Max Node Degree
60
40
20
0
day 12
day 19
day 40
day 54
Embryonic Stem Cells Evolution
RAI1
NRXN1
SLC52A2
BAZ1B
ABR
CNTNAP2
CLIP2
FLII
TBC1D24
GTF2I
SHANK3
DEAF1
BCL11A
PCDH9
SETBP1
PTEN
LIMK1
FOXP2
EBP
ELN
SPTBN4
FDXR
PTCRA
SLITRK6
DPYD
CNTN6
GJC2
ERF
VSX1
TMTC2
SH3TC2
STRC
PCDH15
GJB2
MYH3
CHRNA7
MLXIPL
DIAPH3
ERCC8
TMPRSS3
CNTN5
IQSEC2
EDNRB
OTOF
TIMM8A
KCNA2
SOX10
RFC2
AIFM1
SLC26A4
PEX1
HGSNAT
MORC2
BHLHE22
TBL2
HEXA
TMEM126A
MTMR2
ERCC6
SGPL1
GTF2IRD1
GLB1
ATP1A3
OPA1
MPZ
NDRG1
GM2A
CHRNA4
GIPC3
MYO7A
F11
GRIN2A


A
Demographics
FT ASDf
FT ASDm
FT noASDf
FT noASDm
PT ASDf
PT ASDm
PT noASDf
PT noASDm
BODY WEIGHT(g)
EGA(weeks)
B
No ASD Babies
BW vs. EGA
ExpFit
F
M
BW (grams)
EGA (weeks)
C
ASD Babies
BW vs. EGA
Exp Fit
F
M
BW (grams)
EGA (weeks)
D
Pre Term Babies
Full Term Babies
No ASD Babies
ASD Babies
Frequency
Females
Males


A
Zooming into fluctuations
ABR@70dB μVolts
Frames @25kHz
B
NICU PT ASD A36447L1
ABR@70dB μVolts
Frames @25kHz
C
NICU PT ASD A36447L1
70dB
75dB
80dB
ABR@70,75,80dB μVolts
Frames @25kHz
D
Zooming into 1 Peak
Wave Peak
Amplitude
Width
Prominence
ABR@70dB μVolts
Frames @25kHz
E
Zooming into the Peak
ABR@70dB μVolts
Frames @25kHz
F
$MMS = \frac{Wave^{Peak}}{Wave^{Peak} + Avrg_{min\,to\,min}}$
Micro Movement Spikes
Frames @25kHz