Title: Electrophysiologically-defined excitation-inhibition autism neurosubtypes

Authors: Natasha Bertelsen[1,2], Gabriele Mancini[3,4,5], David Sastre-Yagüe[2,5], Andrea Vitale[1], Gabriel Matías Lorenz[3,4,5], Simone Blanco Malerba[3], Dimitris Bolis[1], Veronica Mandelli[1,2], Pablo Martínez-Cañada[6,7], Alessandro Gozzi[5]*, Stefano Panzeri[3]*,& Michael V. Lombardo[1]*

Affiliations:

1 Laboratory for Autism and Neurodevelopmental Disorders, Center for Neuroscience and Cognitive Systems, Istituto Italiano di Tecnologia, Rovereto, Italy

2 Center for Mind/Brain Sciences, University of Trento, Rovereto, Italy

3 Institute for Neural Information Processing, Center for Molecular Neurobiology (ZMNH), University Medical Center Hamburg-Eppendorf (UKE), Hamburg, Germany

4 Department of Pharmacy and Biotechnology, University of Bologna, Bologna, Italy

5 Functional Neuroimaging Laboratory, Center for Neuroscience and Cognitive Systems, Istituto Italiano di Tecnologia, Rovereto, Italy

6 Department of Computer Engineering, Automation and Robotics, University of Granada, Granada, Spain

7 Research Centre for Information and Communications Technologies (CITIC), University of Granada, Granada, Spain

* Equal senior author contributions
Corresponding author: Michael V. Lombardo (michael.lombardo@iit.it)

Abstract

Excitation-inhibition (E:I) imbalance has long been considered one of the primary neurobiological theories explaining autism. However, the theory has been bolstered heavily by research on high-penetrance genetic mechanisms which are only found in a small proportion of the autism population and can manifest with potentially different directionality. How well does E:I imbalance explain idiopathic autistic males and does such E:I imbalance manifest in one particular direction? Answering this question in human patients necessitates a need for validated tools that allow for inference about E:I balance from non-invasively measured in-vivo electrophysiological (EEG, LFP) time-series data. Predictions from *in-silico* modelling alongside *in-vivo* validations from mouse electrophysiological data show that alterations of synaptic E:I balance cause changes in fractal long memory characteristics of the neural time-series that are captured with a metric known as the Hurst exponent (H). H estimated in resting state EEG data of idiopathic autistic males allows us to discover two distinct E:I 'neurosubtypes' that generalize at high levels (>92% accuracy) in independent data. Each autism neurosubtype captures approximately half of the sample and each can be described by opposing patterns of E:I imbalance relative to a typically-developing (TD) comparison group. Autism E:I neurosubtypes also show differential relationships between H and behavioral and demographic variables consisting of age, intelligence, and autism symptomatology. This work establishes causal evidence for interpreting changes in EEG-derived H as an E:I-relevant biomarker and suggests that idiopathic autistic males can be characterized by opposing types of E:I imbalance with differential phenotypic relevance.

Introduction

Imbalance between neurophysiological excitation and inhibition (henceforth referred to as 'E:I' imbalance) has long been considered one of the primary neurobiological explanations behind autism[1]. While the theory has been highly influential, it still rests on evidence primarily gleaned from animal model research on known high-impact genomic mechanisms linked to autism[2,3]. Nearly all of these genomic causes are rare in frequency, both within the general population but also within autism[4]. Furthermore, the directionality (increased versus decreased E:I balance) and interpretation (initial cause versus homeostatic compensation)[2,3,5] of observed E:I imbalance in such rare genomic causes can markedly differ. Thus, a clear roadblock exists with regard to generalizing heterogeneous evidence of E:I imbalance from such rare genomic causes to autistic individuals whose genetic etiologies are otherwise unknown (i.e. idiopathic autism). Sohal and Rubenstein[3] proposed that the next steps in moving forward are towards identifying more individualized biomarkers of different types of E:I imbalance which might lead towards differential explanations behind heterogeneity in behavioral phenotypes and/or differential responses to treatment. In this work, we take these next steps by assessing whether the population of idiopathic autistic males can be split into E:I neurosubtypes based on *in-vivo* electroencephalography (EEG) biomarkers of high-relevance to underlying E:I balance.

Derivation of biomarkers from neural time-series data such as EEG has gathered considerable attention in recent years[6,7], particularly for its potential impact as a non-invasive *in-vivo* method that could be used with human patients and under important clinical contexts (e.g., treatment, clinical trials)[8,9]. In this work we perform rigorous *in-silico* and *in-vivo* testing and validation of EEG-derived biomarkers that can be sensitive in parsing heterogeneity in E:I balance in idiopathic autistic males. Setting the table for this work, Gao and colleagues[10] observed that the aperiodic signal variation (e.g., 1/f slope) measured from simulated local field potential (LFP) signal from uncoupled excitatory and inhibitory neuronal populations is sensitive to underlying changes in E:I balance, defined as the ratio between the strength of E and I conductances. This effect arises because excitatory AMPA post-synaptic potentials have a faster decay time constant than inhibitory GABA potentials. Thus, an increase of total E over I conductances will result in a relative increase in higher frequency components and consequently flatten the 1/f spectral slope[10]. Our recent *in-silico* modeling has confirmed that such effects are present in critical tests of a more biologically realistic situation when excitatory and inhibitory neuronal populations are recurrently connected[11]. Furthermore, we showed that in addition to metrics like 1/f slope, measures of fractal long-memory characteristics of the time-series such as the Hurst exponent (H) are also sensitive to such effects[11]. We then used causal chemogenetic *in-vivo* manipulations of local E:I balance in mouse prefrontal cortex to experimentally demonstrate that decreases in H measured from BOLD fMRI resting state time series captures causal excitation increases[11]. However, similar causal evidence from *in-silico* and *in-vivo* validations in EEG, and its intracranial counterpart, the local field potentials (LFP), data are still lacking.

It is also important to consider that E:I balance is a complex construct that can be measured at several levels of analysis – from the ratio of E vs I synaptic conductances, to firing rates of E and I populations, and all the way up to MRI-derived phenotypes, such as magnetic resonance spectroscopy (MRS) metabolite concentrations relevant to excitation and inhibition[6]. Gao and colleagues focused on how a biomarker like 1/f slope measured in LFP or EEG might be

specifically linked to the ratio (denoted as *g*) of E vs I synaptic conductances[10]. However, as EEGs and LFP capture many neural parameters[20,24], it is unclear whether such biomarkers are sensitive specifically to synaptic conductance changes, or whether they are instead more sensitive to changes in other E:I-relevant parameters that may covary with synaptic conductances in strongly coupled recurrently connected cortical networks[12,13]. Thus, more work is needed to determine how specific or generalized are such *in-vivo* EEG or LFP biomarkers for inferring specific changes to E:I-relevant parameters such as *g* and/or firing rates. Finally, several types of biomarkers have been proposed in the literature (e.g., H, 1/f slope, total broadband spectral power)[10,11,14], but it is unclear which if any, can best track changes in underlying E:I balance.

Here we start by addressing these mechanistic and explanatory gaps in decoding E:I balance from *in-vivo* measured electrophysiological (EEG, LFP) biomarkers (e.g., H, 1/f slope, total broadband spectral power). Utilizing *in-silico* modeling of recurrently connected E and I neuronal populations, we systematically manipulate a variety of ground truth E:I parameters (e.g., *g*, resting membrane potential) and examine how such changes explain variation in several measured EEG and LFP biomarkers (e.g., H, 1/f slope, total broadband spectral power). *In-silico* modeling predictions are then tested and validated *in-vivo* via causal chemogenetic manipulations of E:I balance in mice. Insights from *in-silico* and *in-vivo* experiments allowed us to scale up this work towards application on resting state EEG data measured in male idiopathic autistic patients. Resting state EEG E:I biomarkers were used to split idiopathic autism into discrete neurosubtypes[15] and then tested for generalizability and differential brain-behavioral relationships.

Results

In-silico evidence that increasing network excitability causes changes in H measured in LFP and EEG

To first test that EEG biomarkers are relevant for inferring underlying E:I balance, we simulated realistic LFP and EEG proxies from a model of a local cortical network of recurrently connected leaky-integrate-and-fire (LIF) point neurons[16] (Figure 1A-D), while manipulating neural parameters related to excitability. This model reproduces well the firing regimes and oscillation spectra of real cortical local networks[17–20]. In particular, the model generates highly realistic gamma (30-100 Hz) oscillations and spectral 1/f slopes. Throughout all simulations we hold the constant input to the model (i.e. $\nu 0$) to a rate of 2 spikes per second per neuron. This level of input results in an asynchronous irregular (AI) network state[16,17] corresponding to output spiking that resembles resting state conditions (Figure 1D). We first varied the ground truth E:I ratio (*g*) – operationalized as the ratio of excitatory AMPA (g_E) divided by inhibitory GABA (g_I) conductances onto the excitatory population ($g = g_E/g_I$) (Figure 1A) - across a wide range that corresponds to around 5.7-14 times greater I than E and covers a range used to model resting state and stimulus-evoked neural cortical activity[11,18,19]. Increasing *g* led to a corresponding increase in the mean firing rate of excitatory neurons (Figure 1E left; Supplementary Table 1) and thus enhanced excitability of the network. However, because E and I are tightly coupled, it also led to an increase on the mean firing rate of I neurons (Figure 1E left; Supplementary Table 1). Measuring H from the simulated LFP and EEG, we found that increasing *g* (i.e. more recurrent synaptic excitation) and consequently, increasing firing rates, leads to decreased H (Figure 1F left;

Supplementary Table 1). Thus, H measured from EEGs or LFP is predictive of two aspects of underlying E:I balance - the ratio of E:I synaptic conductance and the total level of firing in the coupled E:I network.

Because firing rates and the synaptic conductance ratio (*g*) were highly correlated in the above simulations, it is unclear which of the two aspects is more tightly captured by H. This cannot be determined by the uncoupled E:I model of Gao and colleagues[10]. While their work focused on the effect of the ratio *g* of E:I conductances[10], in the uncoupled model of Gao and colleagues similar changes in H (and also in 1/f slope which captures scale-free properties similar to the fractal properties captured by H) occur when the E:I ratio *g* is held constant, but the firing rate of E and I neurons is changed (Supplementary Figure 1). In the uncoupled model, a decrease of H or the flattening of the 1/f slope could be due to an increase in the ratio of E:I conductance without changes in the firing rates, or to an increase in E firing rates over I firing rates without changes in ratio *g* of E:I conductance.

To better understand how H may change in relation to *g* or total firing rate, we thus performed another manipulation of the coupled E:I network to mimic the effect of chemogenetic stimulation of pyramidal neurons via CamkIIα-promoter driven hM3Dq expression[21] (to be used experimentally *in-vivo* in the next section). In this simulated manipulation, we increased the excitability and firing of the E neurons without altering the synaptic conductance *g* ratio. This has the effect of increasing the rates of both E and I neurons without changing the synaptic conductance (Figure 1E center; Supplementary Table 1). In this simulation we found that H also tracks with firing rate change, demonstrating that it is not invariant to changes of simulation parameters as it would be predicted if it only associated with *g*. Moreover, we found that H also tracked with the firing rate changes of a third set of simulations of the coupled E:I network, simulating the effect of chemogenetic inhibition of both excitatory and inhibitory neurons via hM4Di[21] expression under the control of hSyn-promoter[22]. This manipulation (which we tested experimentally, described below) decreases the excitability and firing of both E and I neurons while at the same time decreasing the synaptic strengths of both E and I synapses[22] (Figure 1E right; Supplementary Table 1).

Three main biomarkers of E:I balance have been proposed based on spectral properties of LFP and EEG: the aperiodic 1/f slope[10], the Hurst exponent (H)[11], and the total broadband spectral power[14]. To understand how each relates to changes in synaptic conductance ratios or in firing rates, we also computed the 1/f slope and total broadband spectral power from the same EEG and LFP simulations as above. We found that both the 1/f slope and total broadband spectral power increased with the total firing rate, in all cases, even when the ground-truth synaptic conductance ratio g was unchanged (Supplementary Figure 2; Supplementary Table 1). Thus, based on this model all the three biomarkers correlate with the total firing rate and could be used to predict it. To quantitatively evaluate how well each of these biomarkers predict the total firing rate, we trained a linear regression model to predict the firing rate from the biomarkers measured in a first 'training' set of simulations and we then used the linear model to predict the firing rate from the biomarkers in a second set of independently generated 'validation' simulations. As expected, all three biomarkers linearly predicted the firing rate well (Figure 1G; Supplementary Figure 2; Supplementary Table 1).

In sum, extending our prior observations made on simulated LFP and fMRI BOLD signal[11], we determined that biomarkers (H, 1/f slope or broadband power) measured in simulated aggregate electrical potentials (such as the LFP or EEG) of coupled local cortical networks captures well and primarily the level of firing rates in the network, an important aspect of E:I balance. These indices reflect the operating point (or excitability) of the network in terms of overall firing rate, more than the ratio of E:I synaptic conductances as previously interpreted[10,11].

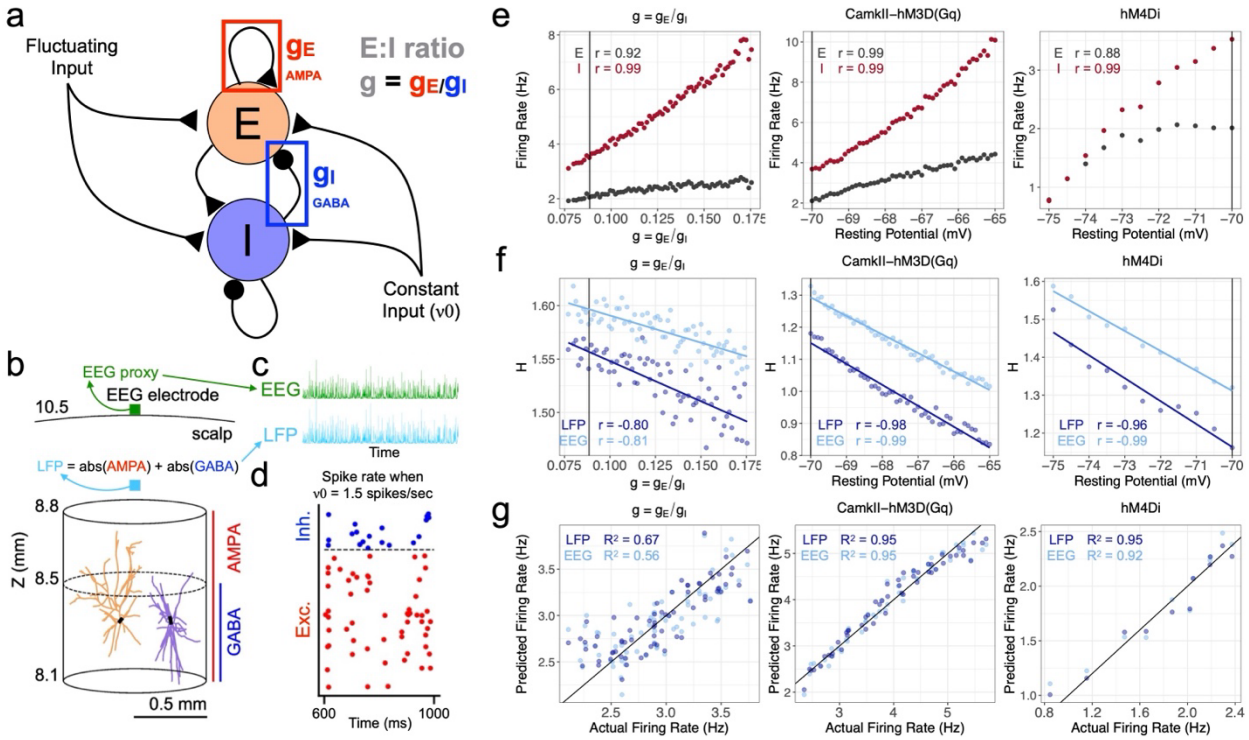


Figure 1: In-silico modeling shows that H computed from LFP and EEG data tracks with underlying changes to network excitability. Panel a *shows a schematic of the recurrently connected leaky-integrate-and-fire point neuron model used to simulate LFP and EEG signal while manipulating ground truth E:I ratio. The model consists of 4000 excitatory (node E,), and inhibitory) neurons (node I). Each E and I population has recurrent connections between and within each population. The E:I ratio is operationalized as g and is derived as the ratio between excitatory AMPA (g_E; red) and inhibitory GABA (g_I; blue) synaptic conductance received by E neurons. The model was simulated over a range of E:I ratios that have inhibition 5.7-14 times greater than excitation and whereby the constant input to the model (v0) is set to 2 spikes/second to resemble resting state conditions.* ***Panel b*** *shows a schematic of a multicompartment neuron model used in prior work to derive an EEG proxy that capture >90% of the variance in ground truth LFP signal.* ***Panel c*** *shows examples of LFP and EEG proxies simulated from the model.* ***Panel d*** *shows an example raster plot of the spiking activity of the model, which was asynchronous irregular.* ***Panels e-g*** *show the results of simulations where g (left) or resting membrane potential are varied either to increase (center) or decrease (right) excitability. The center and right columns are simulations that are meant to reflect changes that occur to excitability of E (center) or E and I (right) neurons as a result of the CamkII-hM3D(Gq) or hM4Di DREADD manipulations that we implemented in subsequent in-vivo validation experiments. The simulated decreases/increases in excitability of the neurons due to these chemogenetic manipulations are primarily modeled as*

decreases/increases of the resting potential of these neurons as indicated in the x axis labels (see Methods). ***Panel e*** *shows how mean firing rate (y-axis) of E (gray) and I (dark red) neuronal populations co-vary as a function of the manipulated parameters (x-axis).* ***Panel f*** *shows how the Hurst exponent (H) computed from LFP (dark blue) and EEG (light blue) data co-vary with changes to the simulated parameters (x-axis).* ***Panel g*** *plots actual ground truth firing rate of a validation set (x-axis) against predicted firing rate from a linear model trained based on H computed from LFP (dark blue) or EEG (light blue) data. Perfect predictions are shown by the black line, while the best fit lines are shown in light and dark blue.*

Experimental animal model evidence that chemogenetically altering E:I ratio in-vivo causes predicted changes to LFP H

The *in-silico* modeling provides a basis for predicting that *in-vivo* upregulation of excitability of excitatory neurons to increase the levels of firing should cause biomarkers measured in electrophysiological extracellular potentials to change in specific directions (e.g. decreased H, increased 1/f slope and broadband power). Here we test these causal predictions with a mouse model whereby we chemogenetically upregulated excitation of pyramidal neurons via overexpression of hM3Dq DREADD receptor under the control of CamkII-promoter in the medial prefrontal cortex (PFC) (Figure 2A-B). We performed electrophysiological recordings multi-unit activity (MUA) and LFPs in the chemogenetically manipulated area. We first verified that the CamkII-hM3D(Gq) DREADD manipulation increased PFC MUA spiking activity (a measure of total local firing[20,23]). Compared to the SHAM control injection, CamkII-hM3D(Gq) activation via systemic administration of DREADD activator CNO significantly and steadily increased MUA levels (by 2-4 standard deviations relative to baseline) over the examined 60 min time-window (Figure 2D; full statistics reported in Supplementary Table 2). Evidence of increased levels of spiking activity could be correctly inferred using H measured from the PFC LFP (Figure 2E; Supplementary Table 2). In the treatment phase, increased excitation caused statistically significant decreases in H of around 2-4 standard deviations relative to baseline in the DREADD condition. This effect was markedly greater than the negligible change in H relative to baseline observed in the SHAM control condition. This causal effect to decrease LFP-derived H was increasingly evident towards the end of the transition phase and continued throughout the treatment phase (Figure 2E; Supplementary Table 2). Thus, these results confirm the *in-silico* modeling predictions that increased excitability and causes a subsequent decrease in electrophysiologically measured H.

We next tested *in-silico* predictions regarding whether downregulating excitability of E and I neuronal populations would cause an increase in H. To experimentally test this hypothesis, we virally expressed DREADD hM4Di receptors in mouse PFC under the control of hSyn promoter. This manipulation has been shown to silence both DREADD-expressing excitatory and inhibitory neuronal populations[21,22] and should thus collectively produce a reduction of MUA activity in the manipulated region. In keeping with this, MUA levels decreased upon activation of hM4Di receptors via systemic activation of CNO (Figure 2F; Supplementary Table 2). Importantly, we also found that the resulting decreased *in-vivo* firing rate levels could be inferred by the increase in LFP-derived H, as predicted by the *in-silico* local network model. In the treatment phase of the experiment hM4Di activation by CNO produced a large increase of H around 3-4 standard deviations relative to baseline (Figure 2G; Supplementary Table 2).

For both types of DREADD manipulations, the 1/f slope was predictive of the manipulation as expected by our *in-silico* model, but it was far less predictive of the manipulation compared to H (Supplementary Figure 3; Supplementary Table 2). Importantly, the result for total broadband spectral power was in the opposite direction compared to what is predicted by our *in-silico* local network model (Supplementary Figure 3). Our *in-silico* model of coupled E:I populations captures fully the effect of local interactions, which generate oscillations within the range of 30-100Hz[16–18]. According to this model, if only local neural interactions were at play, the total broadband power would be an excellent predictor of the firing rate (Supplementary Figure 2). However, in real cortical LFPs, the low-frequency (< 4 Hz) power reflects global network fluctuations rather than local E:I interactions[24]. On our data, as shown also in Rocchi et al.,[22] when the local firing rate decreases or increases, these global components of activity become more or less prominent in the spectra (Supplementary Figure 4). This phenomenon then biases the broadband power metric in the direction opposite to that induced by the variations of local firing rates.

The comparison of SHAM vs DREADD has the advantage that it provides causal evidence that increases or decreases in firing rates can be detected by decreases or increases in H. However, it may be argued that the large variations in network configuration provided by the manipulation are unlike the variations naturally occurring across time or across animals. To evaluate whether H can track naturally occurring variations in firing rate, we compared, separately within the two SHAM and the two (hM3D(Gq) and hM4Di) DREADD conditions, the naturally occurring variations of firing rates with the variations in H during the treatment phase of the experiment. We found a significant negative correlation between firing rate and H, implying that H can track naturally spontaneously occurring as well as drug-induced variations of firing rate (Fig 2F, 2I; Supplementary Table 3). Importantly, on the same data the 1/f slope and the total broadband power showed non-significant and/or inconsistent (that is, with variable sign across datasets) correlation with firing rates compared to H (Supplementary Figure 3; Supplementary Table 3). The correlations with 1/f slope, when present, were positive in sign as predicted by the *in-silico* model. The correlations with total broadband spectral power, when present, were negative in sign, which is consistent with the chemogenetic manipulation results and was the opposite of what was predicted by the *in-silico* model of the local recurrent E:I network.

In sum, our results show that in mouse PFC, H is the only electrophysiologically measured biomarker that can successfully track with both positive and negative variation in firing rates either causally induced with CamkII-hM3D(Gq) or hM4Di DREADD or spontaneously occurring across time or animals. Combined with the evidence from the *in-silico* modeling, this work shows that H measured in LFP and EEG is sensitive to underlying up- or down-regulations in firing rates, an important aspect of E:I balance.

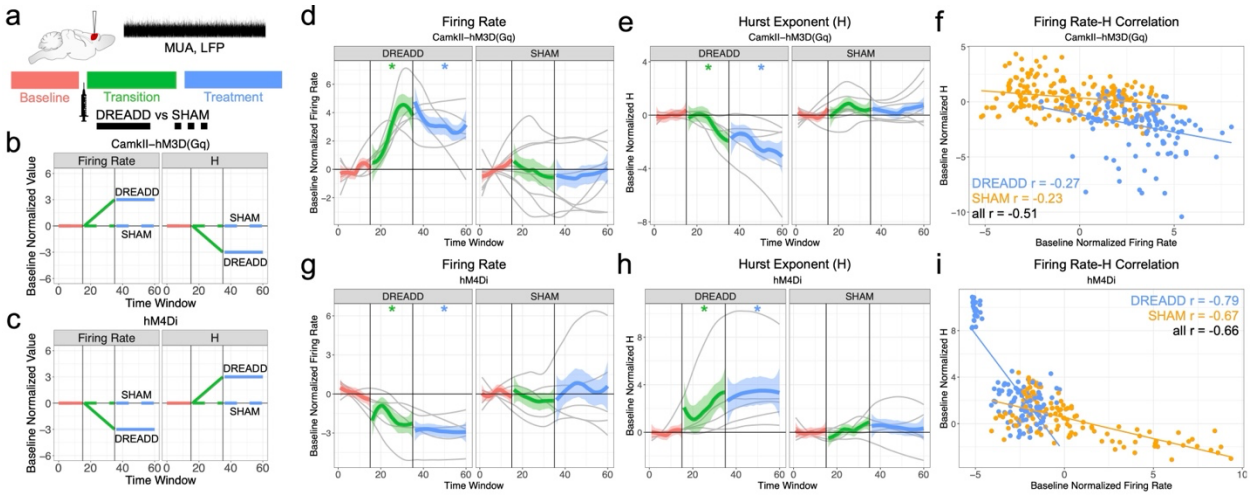


Figure 2: Chemogenetic manipulation of E:I balance causes predicted changes to H. Panel a *shows a schematic of the experimental design of the DREADD excitation (CamkII-hM3D(Gq)) and silencing (hM4Di) experiments. The first 15 minutes of the experiment was considered the 'Baseline' phase (pink) before any injection (DREADD or SHAM) was given. After the 'Baseline', the experimental DREADD or SHAM control injection was given and then a period of 20 minutes ensued, which is dubbed the 'Transition' phase (green) of the experiment, where the DREADD drug is predicted to begin taking effect. After the 'Transition' phase is the 'Treatment' phase (blue), which lasts up until 60 minutes after experiment start and is the phase where the DREADD drug is predicted to have maximal effect. Both multi-unit arrays (MUA) and LFP were measured from the injection site in PFC to measure electrophysiological effects.* ***Panel b*** *shows the predictions for firing rate (left) and H (right) for the DREADD excitation experiment (experimental DREADD condition shown as the solid lines, control SHAM condition shown as the dashed lines).* ***Panel c*** *shows the predictions for firing rate (left) and H (right) for the DREADD silencing experiment (experimental DREADD condition shown as the solid lines, control SHAM condition shown as the dashed lines).* ***Panels d-i*** *summarize the effects of the DREADD excitation (d, e, f) and silencing (g, h, i) experiments.* ***Panels d, e and g, h*** *depict baseline normalized firing rate or H (y-axis) trajectories over sliding time windows (x-axis) whereby with each time window representing a 4 second segment of time and with each window separated by around 1 minute. The y-axis of panels d and g show the baseline normalized firing rates as measured by MUA. The y-axis of panels e and h show the baseline normalized H responses as measured from LFP. All baseline normalized values on the y-axis represent z-scores, and thus, the units on the y-axis in each plot represent units of standard deviation difference from the baseline phase and 0 represents the baseline average value. The experimental phases are denoted with color (baseline, pink; transition, green; treatment, blue). Each plot's facets represent either the experimental (DREADD, left) or control (SHAM, right) conditions. The gray lines in each plot represent the data for each individual mouse, while the thicker colored lines represent the group fit with confidence bands. All trajectories (individual and group) are fitted with a LOESS kernel. The vertical black lines in represent the start of the transition and treatment phases respectively, while the horizontal line at 0 visually shows where the baseline average is for each plot.* ***Panels f and i*** *show associations between baseline normalized firing rate (x-axis) and baseline normalized H (y-axis) from trials in the treatment phase of the CamkII-hM3D(Gq) DREADD excitation (F) or hM4Di DREADD silencing (I) experiments and with DREADD and SHAM conditions indicated by pink and turquoise colors respectively.*

Identifying EEG-derived autism E:I neurosubtypes

Having established causal *in-silico* and *in-vivo* evidence that H measured in LFP and EEG is driven by changes in underlying E:I balance, we next translationally extended this work to humans and tested whether the population of male idiopathic autism can be split into EEG-derived E:I 'neurosubtypes'. Here we applied unsupervised stability-based relative clustering validation analysis (*reval*)[25] to a relatively large EEG dataset of male autistic individuals (n=286) 5-21 years of age from data releases 1-10 from the Child Mind Institute Healthy Brain Network (CMI-HBN) dataset[26]. Genetic data was not available for analysis on this sample, and thus our usage of the term 'idiopathic' refers to the lack of knowledge regarding known genetic causes of autism. EEG data was examined under the two resting state conditions - eyes open and eyes closed. H was estimated across 93 electrodes across the scalp and then input into the *reval* clustering pipeline.

For both eyes open and eyes closed conditions, we find that the optimal number of subtypes is 2 (Figure 3A-B). Generalization accuracy for this 2-subtype solution in the held-out validation set is 92% for the resting state eyes open condition and 98% in the resting state eyes closed condition. This 2-cluster solution is indicative of true clusters, as it heavily deviates from a single multivariate Gaussian null distribution (SigClust analysis: all $p < 0.0015$ in training and validation sets for eyes open and eyes closed) (see Supplementary Table 4). Plots of dimensionality-reduced data with UMAP further show visually how the subtypes each have their own distribution and distinct peaks (Figure 3A-B). Thus, using EEG H patterns across the entire scalp, unsupervised data-driven clustering can robustly and replicably identify autism E:I 'neurosubtypes' that generalize with high-accuracy in new unseen datasets. These findings solidify the idea that at the very least, autism can be split into at least two E:I-defined neurosubtypes based on EEG biomarkers such as H.

In terms of the proportion of individuals within each subtype, we find that these E:I neurosubtypes comprise an approximate half-split of the entire autism sample examined (eyes open: autism subtype 1 = 45.45%, autism subtype 2 = 54.55%; eyes closed: autism subtype 1 = 52.10%, autism subtype 2 = 47.90%). A vast majority (n = 211; 73.78%) of participants are clustered into the same subtype across the two eye conditions (Figure 3C). The smaller subset of participants who changed subtypes between eyes open and eyes closed (n = 75; 26.22%), were not different from those that stayed in the same subtype in terms of age, FIQ, or autism symptomatology measures (Supplementary Table 5).

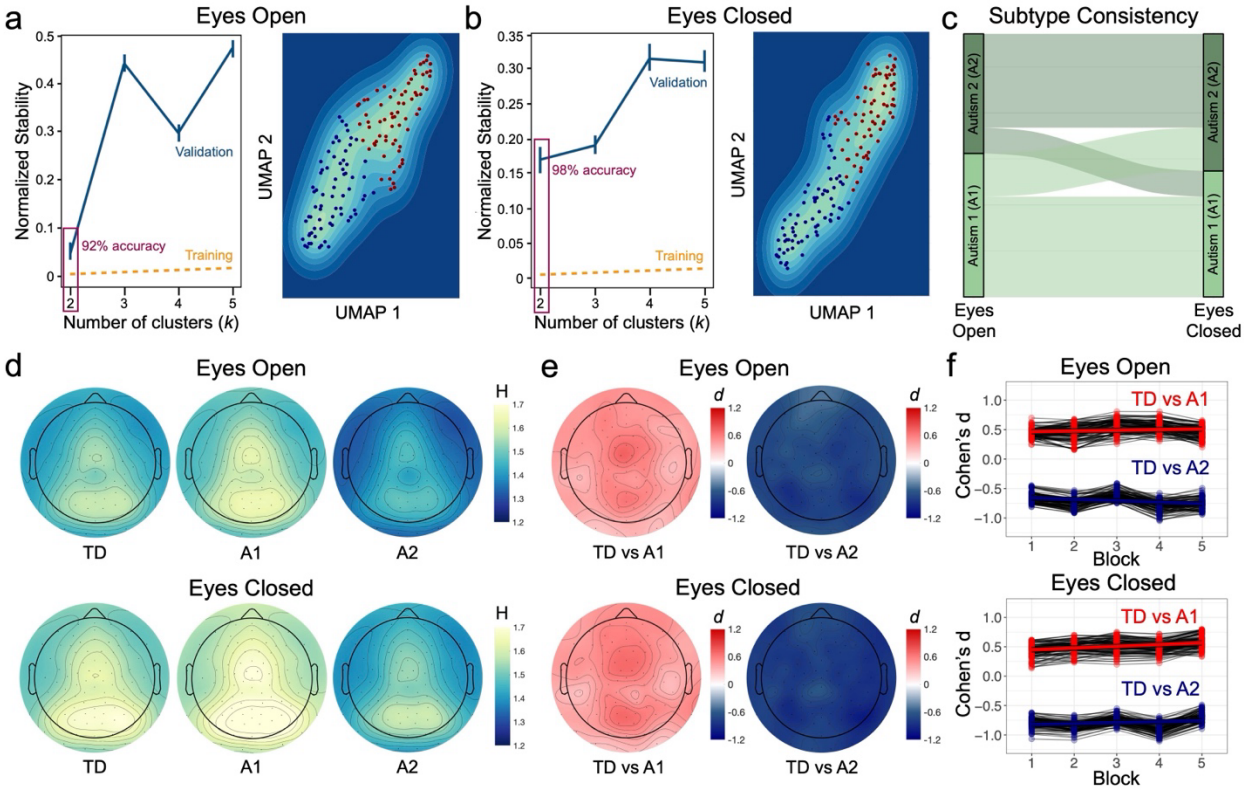


Figure 3: Identification of autism subtypes with high generalizability accuracy based on EEG-derived H. Panels a-b *show normalized stability plots (eyes open, a; eyes closed, b) that indicate that a 2-cluster solution minimizes clustering stability and produces high generalizability accuracy (92-98%) in independent data. The UMAP plots next to the stability plots show evidence that the clusters highly deviate from the null single multivariate Gaussian distribution.* ***Panel c*** *shows an alluvial plot to indicate how consistent are the 2 subtypes across eyes open and eyes closed conditions.* ***Panel d*** *(left) shows block-averaged H topographies for eyes open (top) and eyes closed (bottom) conditions. After statistical testing, we quantified the effect size difference between autism subtypes and TD. These effect size topographies are shown in* ***panel e****, whereby we see that autism subtype 1 has ubiquitously increased H across all electrodes (left), whereas autism subtype 2 has ubiquitously decreased H across all electrodes (right).* ***Panel f*** *shows the effect sizes plotted for all electrodes and across each of the 5 blocks. These plots show that the effect sizes remain relatively consistent over 5 repeat blocks of eyes open or closed conditions.*

Autism E:I neurosubtypes differ from typical development in opposite directions

We next proceeded to test for differences between autism E:I neurosubtypes relative to a typically-developing (TD) comparison group. Linear mixed effect models were used with H values at each electrode as the dependent variable. Subtype, age, and their interaction were fixed effects in the model, while acquisition site and repeat blocks were modeled as random effects (see Methods). We found a significant main effect of subtype across all electrodes. Specific between-group comparisons of TD vs autism subtype 1 and TD vs autism subtype 2 show that the group difference here is driven by both autism subtypes being significantly different from TD. Autism subtype 1 had higher H compared to TD (eyes open condition: *Cohen's d* > 0.27, *mean d* = 0.49; eyes closed condition: *Cohen's d* > 0.25, *mean d* = 0.51) (Figure 3D-E; Supplementary Table 5).

While the group differences are relatively strong across all electrodes, the topographies of effect size for TD vs autism subtype 1 shown in Figure 3E indicate relatively larger differences over central and posterior electrodes. In contrast, autism subtype 2 showed the opposite effect of lower H compared to TD (eyes open condition: *Cohen's d* < -0.52; *mean d* = -0.71; eyes closed condition: *Cohen's d* < -0.62, *mean d* = -0.79) (Figure 3D-E). These effects were all remarkably consistent over repeated blocks, which further emphasizes the robustness of the effects (Figure 3F). While group differences are relatively strong across all electrodes, the topographies of effect size for TD vs autism subtype 2 shown in Figure 3E indicate relatively larger differences over temporo-parietal electrodes. Lower H in autism subtype 2 is indicative of a higher excitability and is compatible with the predictions from the original E:I theory[1]. However, the existence of the opposing autism subtype 1 with higher H, is compatible with a more updated view about E:I imbalance in autism which suggests that imbalance towards either extreme may be important[2,3]. The subtypes are not equally deviant compared to the TD group though, as autism subtype 2 tends to show a more extreme deviation than autism subtype 1. In sum, the evidence here showcases an important validation that idiopathic autistic males can be described by E:I imbalances that deviate from the TD population in either direction.

In addition to main effects of subtype, we also identified ubiquitous main effects of age involving a majority of all electrodes in the eyes open condition and primarily central and occipital electrodes in the eyes closed condition (Supplementary Figure 3A-B; Supplementary Table 6). These effects are described as a negative relationship between H and age (Supplementary Figure 3C-D) and are congruent with other literature on E:I-sensitive metrics (e.g., 1/f slope) and how they change in a similar direction as one ages[27–29]. Finally, given the lack of any significant subtype*age interactions, it is clear that these age effects do not manifest differentially as a function of subtype membership (Supplementary Table 6).

Given that statistically significant main effects of subtype were present across the scalp, this potentially indicates that H variance manifests as a global factor. To better understand whether differential subtype H variance is best explained by this global or other local factors, we next proceeded to use principal components analysis (PCA) to decompose scalp-wide H variance into orthogonal components with different spatial topographies. PC1 explains a large majority of scalp-wide H variance (eyes open = 67.1%; eyes closed = 66.6%) and is spatially homogeneous across the scalp. This homogenous topography is indicative of PC1 being a global component. Three other PCs were also identified, and cumulatively with PC1, account for ~80% of the total variance in H (Supplementary Figure 4). These subsequent PCs show a much more specific and localized spatial topography. Thus, beyond the global H signal captured by the large PC1 component, there are smaller components explaining how H more subtly varies in a spatially-specific manner.

We next ran similar linear mixed effect models testing subtype, age, and subtype*age interactions on these first 4 PCs (Supplementary Table 7). This was done in order to test in a more principled manner whether the subtypes are differentiated by global and/or local components of H variation. Results for PC1 are similar across both eyes open and eyes closed conditions, with no subtype*age interactions (eyes open: $F = 2.42$, $p_{fdr} = 0.359$; eyes closed: $F = 0.80$, $p_{fdr} = 0.756$), but significant main effects of age (eyes open: $F = 16.32$, $p_{fdr} = 2.73\text{E-}04$; eyes closed: $F = 9.3$, $p_{fdr} = 0.005$) and subtype (eyes open: $F = 144.95$, $p_{fdr} = 3.38\text{E-}46$; eyes closed: $F = 180.38$, $p_{fdr} = 6.48\text{E-}56$). The main effect of age manifests as a negative relationship (i.e. decreasing PC1 scores

with increasing age) (Supplementary Figure 3E-F). The main effect of subtype manifests with autism subtype 1 having higher values compared to TD (eyes open: t = 7.06, p_{fdr} = 6.45E-11, *Cohen's d* = 0.62; eyes closed: t = 6.79, p_{fdr} = 2.91E-10, *Cohen's d* = 0.65), but autism subtype 2 showing lower values compared to TD (eyes open: t = -8.57, p_{fdr} = 3.06E-15, *Cohen's d* = -0.93 ; eyes closed: t = -9.83, p_{fdr} = 6.52E-19, *Cohen's d* = -1.04) (Figure 4A-B). Thus, results from PC1 captures the same types of effects observed in prior analyses per each electrode.

During the eyes open condition, PC4 was the only other component to show a significant main effect of subtype and with significant differences between TD and the autism neurosubtypes (F = 3.53, p = 0.03 p_{fdr} = 0.061; TD vs A1: t = 2.19, p_{fdr} = 0.04; TD vs A2: t = 2.46, p_{fdr} = 0.029; Supplementary Table 7). The topography of PC4 shows differentiation between frontal and temporal electrodes (Figure 4C). Given the combination of PC1 and PC4 subtype effects during the eyes open condition, we next reconstructed the data exclusively based on PC1 and PC4 in order to descriptively show how subtype main effects manifest. Here we see that H is elevated in autism subtype 1 compared to TD and this effect is most prominent over frontal and posterior electrodes compared to temporal electrodes (Figure 4D). In contrast, H is decreased in autism subtype 2 compared to TD, and this effect is strongest over non-frontal electrodes (Figure 4E).

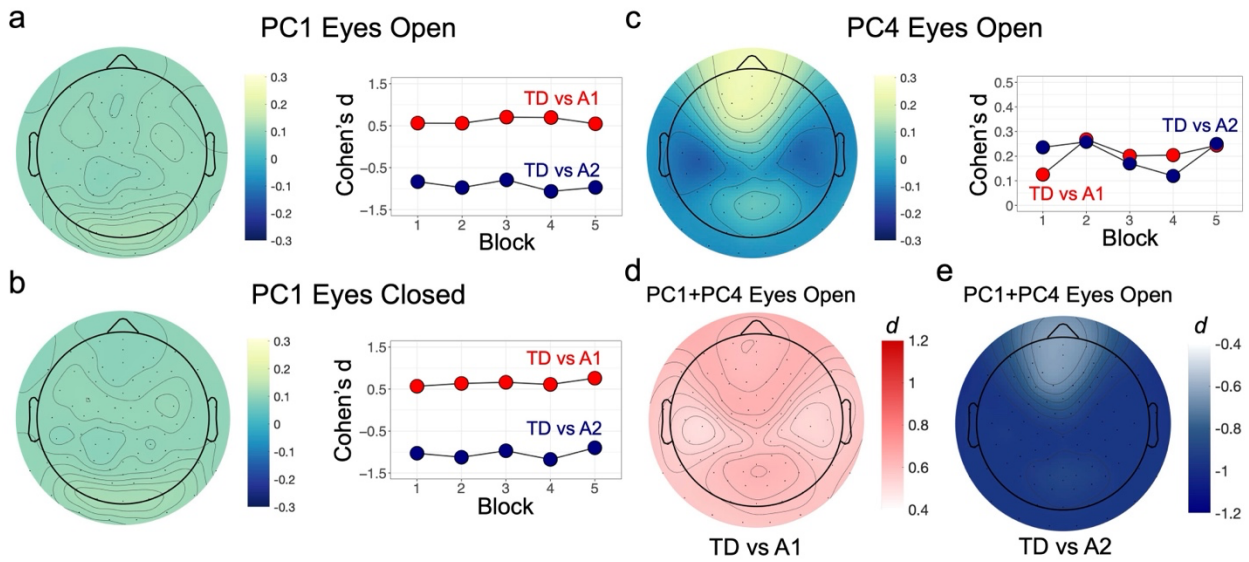


Figure 4: PCA decomposition of global versus local components of H that differ significantly between autism neurosubtypes and TD. Panels a and b *show PC1 for eyes open (top) and eyes closed (bottom) conditions. The topography of PC1 is shown on the left, with similar loadings across all electrodes, indicating that it is a global component explaining a large majority of the H variance. Effect size plots are shown to the right of the topography plots and indicate remarkable consistency of effect size across repeat blocks.* ***Panel c*** *shows the topography and effect sizes for the significant main effect of subtype in PC4 during the eyes open condition. PC4's topography has strong loading over frontal electrodes and opposing effects at over temporal electrodes. Reconstructing PC1 and PC4 back to original H space shows increased H in autism subtype 1 compared to TD that is most prominent over frontal and posterior electrodes relative to temporal electrodes (****panel d****). In contrast, H is decreased in autism subtype 2 relative to TD and this effect is most prominent over non-frontal electrodes (****panel e****).*

Differential brain-behavior relationships

We next sought to test for phenotypic differences between these neurally-defined H autism subtypes in age, FIQ, and autism symptomatology (e.g., ASSQ, SCQ, SRS, RBS). None of these variables showed significant differences between subtypes that were consistent across eyes open and eyes closed conditions (see Supplementary Table 8 for full set of statistical comparisons for all dependent variables) (Supplementary Figure 5). Although on-average phenotypic differences were not strong between the subtypes, it is possible that relationships between H and these phenotypic measures may be different. To test this hypothesis, we used partial least squares (PLS) correlation analysis[30]. This analysis identified one significant latent variable pair for both the eyes open (d = 6.09, p = 3.99e-4, percentage covariance explained = 33.1%) and eyes closed (d = 9.79, p = 2.99e-4, percentage covariance explained = 64.61%) conditions. Effects in both eyes open and eyes closed conditions are likely driven by a global effect spread across the scalp, as PLS brain bootstrap (BSR) ratios are widely greater than 2 across all electrodes (Figure 5A-B). In terms of the behavioral variables that drive these brain-behavior relationships, we see some evidence suggesting that age, SCQ, and SRS are important (Figure 5C-D), primarily in TD and autism subtype 1. Across eyes open and eyes closed, these effects are highly preserved within both autism subtypes (autism subtype 1: r = 0.88, p = 0.001; autism subtype 2: r = 0.84, p = 0.004). However, the effects in the TD group are not as highly preserved across eyes open and eyes closed (r = 0.51, p = 0.15). This relatively weaker preservation across resting state conditions in TD may explain why age, SCQ, and SRS correlations change in the TD group across eyes open or closed conditions. The weaker preservation in TD may also be indicative of higher sensitivity to neural changes of behavioral relevance in TD across relatively simple eye viewing condition changes. In contrast to the variables of importance for TD and autism subtype 1, the only variable with a strong driving effect in autism subtype 2 is age and this effect is consistent across eyes open and eyes closed conditions (Figure 5C-D). The relationship here is negative and implies that with increasing age, H decreases in autism subtype 2.

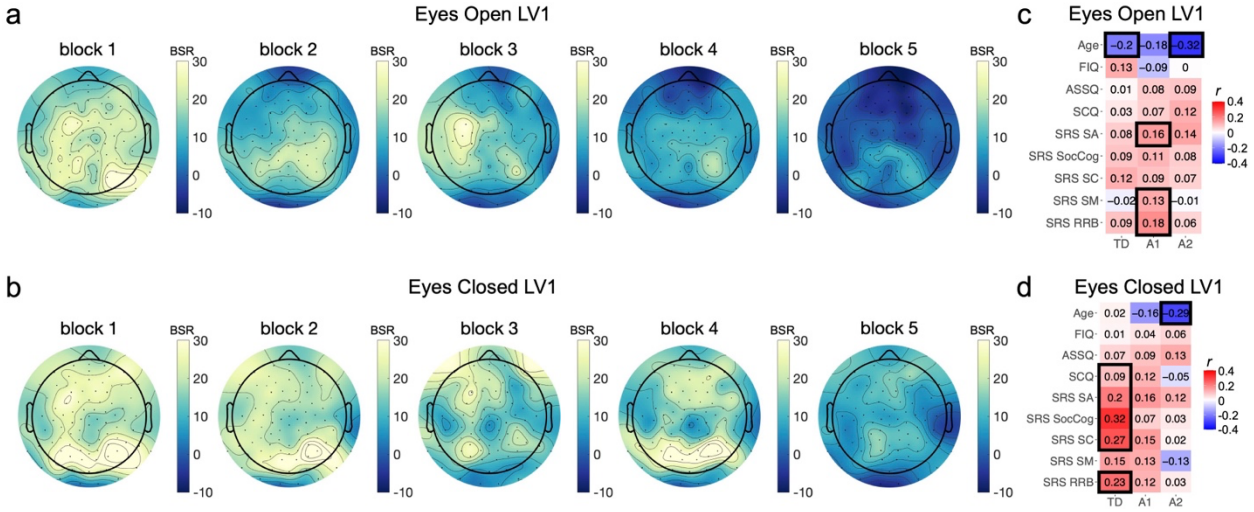


Figure 5: Subtype-specific brain-behavior relationships. *PLS analysis identified one latent variable pair, LV1, within eyes open and eyes closed conditions that shows a statistically significant multivariate brain (H) by behavioral relationship. Panels A and B show brain bootstrap ratios (BSR) from PLS LV1 across all electrodes and all blocks in the eyes open (**panel a**) and eyes closed (**panel b**) conditions. Brain BSRs > 2 are generally considered strong drivers of the PLS*

relationship, and here all electrodes have brain BSRs well over 2. ***Panels c and d*** *show PLS behavior scores (i.e. correlations per each variable) to illustrate strength and stability of relationships driving the PLS latent variable pair LV1. These heatmaps show correlation values with H and cells with a black outline are those with 95% confidence intervals that do not include a value of zero, and thus, are strong and stable contributors to the LV1 relationships observed in eyes open (****panel c****) and eyes closed (****panel d****) conditions.*

Discussion

Recent years have seen a surge of interest regarding the use of neural time-series features as potential markers of E:I mechanisms. Several indices have been proposed and widely used in the literature[10,11,14]. However, several major problems still need to be addressed to be able to use robustly these indices to infer aspects of E:I balance. First, very few studies have validated the effectiveness of these biomarkers with direct causal manipulations[10,11,31]. Most studies have employed systemic pharmacological manipulations, including the use of general anesthetics[10,31], that typically do not offer unambiguous mechanistic interpretations. This is because systemic pharmacological manipulations or the use of anesthetics do not allow to probe if the observed change in spectral features (for instance measured at a specific electrode location) reflect local changes in E:I balance, or instead larger scale effects due to the systemic nature of the intervention. Moreover, general anesthetics typically have unclear mechanism of action, or involve multiple competing contributions that cannot be simply described as changes of E:I ratio. Second, most studies have concentrated on a single proposed biomarker, and systematic comparisons of the effectiveness of different biomarkers on the same dataset are lacking. Third, E:I imbalance is a nuanced phenomenon dependent on and characterized by many neural features, including levels of firing rates and synaptic conductances of E and I neurons. These factors are tangled because of the strong interactions between E and I neurons in the cortex, and it has been difficult to understand which factors are best captured by the biomarkers.

Here we added three significant advances to address these open questions. First, we added a direct causal validation of biomarkers, by considering the direct effect of chemogenetic up- and down-regulation of local firing rates, an important aspect of E:I balance, on the extracellular potential recorded in the manipulated region. We found that large and robust changes to H are directly caused by these up- and down-regulations of this aspect of local E:I balance. We also showed that changes in H are not only predictive of the relatively large changes in firing rates caused by the causal manipulations, but are also predictive of the more subtle moment-to-moment natural spontaneous variations in firing. This is important because it shows that the H obtained from the extracellular potential is sensitive enough to detect relatively small, naturally occurring variations in rate. Importantly, the ability of H to detect these changes in firing rates was supported by independent *in-silico* modelling results that revealed E-I interaction mechanisms underlying the relationship between H and firing rates. Second, we compared systematically three major biomarkers (H, 1/f slope, broadband power) that has been proposed in the literature, and we found that the H was the only one that was revealing of firing rates across all conditions in *in-silico* and *in-vivo* experiments. The inability of total broadband power to accurately track firing rates in *in-vivo* data can be explained by the fact that changes in low-frequency power - a dominant component of the total power - do not necessarily reflect local interactions. In our view, the advantage of H over 1/f slope could hint at the possibility that E:I-relevant signal in the data comes

from both periodic and aperiodic components of the signal. Estimation of aperiodic components of the signal with methods such as FOOOF[28] attempt to remove periodic components of the signal before estimating 1/f slope. In contrast, H does not a priori factor out periodic signal components. In our data there are large variations of oscillation power at high and low frequencies both due to chemogenetic manipulations and natural variations (see Rocchi et al.,[22], Figure 2 and Supplementary Figure 4). Thus, H may better capture the combination of periodic (oscillatory) and aperiodic components of the signal relevant to E:I mechanisms. Alternatively, the removal of oscillatory power needed to compute the 1/f slope may not always be accurate. For example, because oscillatory power at difference frequencies varies across time or conditions, these inaccuracies would make the 1/f estimation noisier. This is compatible with our observations that there are more variable estimates of 1/f obtained from the model compared to the estimates of H from the same data (Supplementary Figure 1). Third, we established by realistic *in-silico* modelling that when E and I are tightly coupled, H and 1/f slope are more predictive of network excitability (as expressed by firing rate levels) rather than as originally propose, of ratios of E and I conductances. Furthermore, these effects obtained with electrophysiological measures of neural activity (e.g., LFP) were more sensitive to such changes than those seen in prior work using fMRI BOLD signal[11]. For example, while this study showed that H measured in LFP is modulated in opposite directions by chemogenetic manipulations that enhance or reduce excitation, H measured in fMRI BOLD signal may be primarily susceptible to enhancing excitation[11]. To inform future study design or for comparing results obtained with different recording modalities, it will be important to corroborate our studies with further E:I manipulations coupled with scalp, rather than intracranial, EEG recordings. Yet, even with the above open question, together these results uphold the validity and high sensitivity of the H biormarker based on electrophysiological recordings, and suggest that changes in electrophysiological H should be interpreted in terms of changes in excitability reflecting firing rate levels.

This mechanistic insight about H allows for clear interpretation of the EEG-defined E:I neurosubtypes identified in the autism population. We find evidence supporting the idea that within idiopathic autistic males, there are 2 robust and reproducible E:I neurosubtypes and that such subtypes can be identified with 92-98% accuracy in new independent data. Alongside our other past work taking a similar relative clustering validation approach[25,32,33], this work showcases that we can begin with unsupervised data-driven discoveries and then immediately translate those discoveries into supervised prediction and classification models. Further utilization of such robust, reproducible, and highly generalizable stratification models may be useful for future work in the field. The robustness of the observed subtypes can also be seen in the remarkable consistency of the subtype labels across eyes open versus eyes closed conditions in a large majority of individuals, and by the consistency of effect sizes across repeated blocks of eyes open and eyes closed conditions.

Regarding the description of the identified autism subtypes, we find that autism subtype 1 is characterized by higher H compared to TD controls, while autism subtype 2 is characterized by lower H compared to TD and which is most prominent over non-frontal electrodes. The current sample was roughly split in half by these subtypes, indicating that one subtype does not necessarily adequately describe all autistic individuals. The most updated version of the E:I theory of autism predicts that E:I imbalance is affected in autism in either direction compared to what is considered an optimal level of balance defined by what is normative in the typically-developing brain[3]. Our

findings provide empirical validation of this prediction by showing that around half of autistic males in our sample conform to either extreme. Thus, like other work examining heterogeneity in E:I mechanisms in autism[11], this work provides further evidence that dysregulation of E:I in autism can take many forms and may need to be personalized with respect to the type of individual in question.

Another observation from the current results is the remarkably global effect of H in these subtypes. While spatial specificity is a limitation behind EEG data, it is remarkable that most of the effects derived in this work are those that are global in nature across the entire scalp, rather than more localized in topography. The subtle exception to this was a small effect specific to the eyes open condition isolated in PC4. PC4 spatially varies in a manner that distinguishes frontal and temporal electrodes. Reconstructing the data based on PC1 and PC4 components shows that the global attenuation of H in autism subtype 2 is most pronounced over non-frontal electrodes. In contrast, the elevation of H in autism subtype 1 is most pronounced over frontal and posterior electrodes. Thus, the subtypes are different not only in the directionality of their difference relative to TD, but also in the topography of effects, potentially indicative of differential impact of separate brain regions. It is unclear how such results may hold up against other past work looking at H or related measures in autism in EEG[34,35] and rsfMRI data[11,36–38]. Such past work does not examine data-driven E:I subtypes and instead rely on the clinical diagnosis to conduct case-control modeling. Future work applying such data-driven subtyping models will allow us to see how well these observations generalize across EEG and fMRI data and across autistic individuals at different points in the lifespan. Given the lack of more precise spatial sensitivity for localizing these observed atypical EEG E:I effects, future work examining these kinds of subtypes with rsfMRI may be important. It could be that E:I mechanisms may be differentially altered along spatial gradients of macroscale neural organization such as sensory-association cortex hierarchy[39], which are also known to be sensitive to differences in intrinsic neural time-scale gradients and E:I mechanisms[40,41].

Finally, we found little evidence of on-average subtype differences in demographic and behavioral measures such as age, IQ, and autism-specific symptomatology. This suggests that E:I neurosubtypes are not immediately self-evident via proxies of dramatically different phenotypic profiles alone. Rather, what is more sensitive to explaining behavioral phenotypic effects are the subtype labels combined with individual-level variation in H. PLS modeling allowed us to identify a significant multivariate brain-behavioral relationship that manifests differently between the 2 autism E:I neurosubtypes. Autistic traits, as measured by the SRS, are positively associated with increasing global H in autism subtype 1 during the eyes open condition. This effect is relatively preserved during the eyes closed condition and also evident in TD. Thus, the effect here may be indicative of a continuum for how E:I biology may be related to autistic traits in TD and autism subtype 1. In contrast, a similar type of relationship was not apparent between SRS and H in autism subtype 2. Given that continuous variation in autistic traits may be mediated by heritable polygenic risk mechanisms[42,43], these results may be suggestive that the differential E:I neurophenotype in autism subtype 1 may be mediated by different neurobiological and genomic mechanisms compared to autism subtype 2. Thus, a potential inference from these results is that E:I autism neurosubtypes may be more enriched for different types of genomic and neurobiological mechanisms. Identification of E:I neurosubtypes with differential brain-behavioral relationships may also be important for translational research to facilitate more positive outcomes for specific

types of individuals. Given prerequisite pre-treatment EEG datasets, current or future clinical trials focusing on pharmacological agents that manipulate/alter E:I-relevant mechanisms could be designed and/or re-analyzed in ways that take into account such stratification biomarkers[6]. Such an approach might yield more powerful and personalized insight into how such E:I-relevant agents may exert an effect (if any) on important clinical outcomes.

There are limitations and caveats to underscore about the current work. First, the autism dataset only examines males. Examination of females was not included due to limited sample size. However, given the potential that E:I mechanisms may not operate the same way in autistic males versus females[11], future work examining females is needed. Second, although the IQ distributions in autism extend into the lower ranges, a majority of autistic individuals falls within 2 standard deviations of the TD norm. Thus, generalizing this work to autistic individuals with very low IQ may be limited. Third, the minimum age in our sample was 5 years and thus, our ability to generalize to even younger individuals may be limited. Given that E:I mechanisms tend to change with age (Supplementary Figure 3)[27–29], more work is needed examining these types of metrics at the earliest ages autism can be diagnosed. The current version of the E:I theory[2,3] allows for the possibility that E:I imbalance may arise because of core pathophysiology versus compensatory effects. As individuals age, there may be higher likelihood that E:I phenotypes may be indicative of compensatory effects. Fourth, future work should examine these subtypes and biomarkers in individuals that are better characterized in terms of genomic mechanisms. This will better elucidate the impact of differential genomic mechanisms (e.g., high impact rare variants versus polygenic risk) on the subtypes. Finally, our work here is restricted to what occurs during resting state eyes open and eyes closed conditions. It is possible that when processing dynamic naturalistic stimuli and/or across specific cognitive tasks or conditions[44,45] different results might arise. Thus, future work examining E:I-relevant biomarkers and stratifications is needed across a range of conditions, stimuli, and contexts (e.g., social interaction) that might better capture the types of situations autistic individuals exhibit strengths and/or weaknesses.

In conclusion, we have discovered that idiopathic autistic males can be split into E:I neurosubtypes based on EEG data collected during resting state conditions. These neurosubtypes exhibit opposing effects in relation to a TD comparison group and are differentially related to phenotypic measures. Future work expanding on these insights may extend the work into females, individuals with more extreme IQ profiles, younger ages, and more comprehensive genomic characterization. Finally, the work could provide a basis for examining how E:I mechanisms may be affected in other cognitive states and/or while processing other types of stimuli.

Methods

CMI-HBN dataset

To identify autism E:I neurosubtypes, we utilized EEG and phenotypic data from data releases 1-10 of the Child Mind Institute Healthy Brain Network (CMI-HBN) dataset. The CMI-HBN is an initiative to create a biobank of individuals aged 5-21 from the New York area. Participants undergo diagnostic and phenotypic assessments, EEG and MRI scanning during resting state, naturalistic viewing and task conditions. Data was acquired at four sites in the New York area - Staten Island, Midtown, Harlem, and a mobile unit. The CMI-HBN initiative received ethical approval by the Chesapeake Institutional Review Board, and written informed consent was obtained from all participants or their legal guardians (for participants < 18 years). For further details on the CMI-HBN initiative see Alexander et al., 2017[26].

For this work we started by isolating all participants from data releases 1-10 that had a diagnosis of autism (n = 412) or who were typically-developing (TD), as indicated by the label 'No Diagnosis Given' (n = 257). From this pool of data, we excluded participants with missing resting state EEG data (n = 15) or whose data could not be successfully preprocessed (n = 106). Given a low sample size of female participants with autism (n = 50) and a potential sex-bias on the Hurst exponent, female participants (n = 139) were also excluded. The final sample size was of n = 286 autistic males and n = 123 typically developing males. The Wechsler Adult Intelligence Scale (WAIS), Wechsler Abbreviated Scale of Intelligence (WASI) and Wechsler Intelligence Scale for Children (WISC) were selected as measures of full-scale IQ (FIQ), while the Social Responsiveness Scale (SRS-2), Repetitive Behavior Scale (RBS-R), Social Communication Questionnaire (SCQ) and Autism Spectrum Screening Questionnaire (ASSQ) were selected as clinical phenotypic measures. See Supplementary Table 9 for a summary of participant characteristics.

EEG data acquisition

EEG data was recorded using a 128-channel EEG geodesic hydrocel system (Electrical Geodesics Inc; EGI). Data was recorded at a sampling rate of 500 Hz, with a bandpass filter of 0.1 to 100 Hz and reference at the vertex of the head (Cz). Resting state data was collected for 5 minutes, while the participant sits in front of a screen with a fixation cross on the center. Throughout the recording, participants alternate having their eyes open (20 seconds blocks) and closed (30 seconds blocks) for five blocks.

EEG data preprocessing

In order to handle in a reproducible way large amount of data, we developed a semi-automated preprocessing pipeline that combines a set of custom MATLAB scripts executable within the EEGLAB[46] framework with other associated software (ASR[47,48] and IClabel[49]). The pipeline consists of a sequence of steps. Step one consists of removal of the outer ring of channels (26 channels) that are mainly located on the face and the neck and generally with a defective contact with the skin surface. Data was then downsampled to 250Hz (one sample every 4ms) and

a basic FIR filter (*filtfilt.m* function) with a high pass cutoff of 1Hz and a low pass cutoff of 80 Hz was applied. Line noise was removed via a notch filter centered at 60 Hz +/-2. Bad channels were identified based on a combination of three different parameters: 1) flatline duration (for more than 5 consecutive seconds); 2) standard deviations (> 4) as compared to the total channel signal; 3) correlation with the neighboring channels (lower than the default value of 0.8). Channels falling into one of these conditions were rejected and interpolated with a spherical method. In a further step the Artifact Subspace Reconstruction (ASR) algorithm was applied to detect data segments contaminated by non-stereotypical artifacts. A preliminary calibration of the rejection threshold criteria is computed on a 'cleaner' portion of the EEG signal. Then ASR identifies the artifact subspace and repairs the artifactual samples based on the predefined threshold values. After re-referencing to the average of all sensors, the number of channels was further reduced by 10% to match the number of available samples input into independent component analysis (ICA with ica_type set to 'runica'). The resulting components were classified by the ICLabel algorithm as deriving from 'brain' or a variety of non-brain sources (e.g., muscle, line noise, channel noise, eye, or other). The non-brain ICs (probability < 20% of being 'brain') were then projected out from the data. After preprocessing was complete, we output preprocessing reports that could be manually checked and classified for exclusion or inclusion into further downstream data analysis. For this data quality control step, all preprocessing reports were manually inspected by MVL and a subset of overlapping subjects were also inspected by NB and AV. From this data quality control check we excluded participants for a variety of qualitative and quantitative reasons, including issues with the raw time-series data, large numbers of bad channels or bad samples that could not be repaired, too few high probability brain independent components (IC) in the top-35 ranked ICs, and/or odd looking IC topographies amongst ICs labeled as 'brain'. For examples of excluded and included participants, see Supplementary Figure 6. Data quality metrics such as number of bad channels and percentage of bad samples were computed for all participants retained for further downstream analysis and were used for further denoising steps before the subtyping analysis.

EEG subtyping analysis

For EEG-based neural features relevant to E:I imbalance, we computed the Hurst exponent (H). Prior work has validated the use of H to infer E:I mechanisms from neural time-series data[11], whereby lower H values are indicative of increased E:I ratio. H was computed at each electrode for each of the five blocks of eyes open or eyes closed conditions with the *bfn_mfin_ml.m* function from the *nonfractal* MATLAB library (https://github.com/wonsang/nonfractal). After computing H for each participant and electrode, we implemented a denoising procedure to remove variance associated with data quality metrics from the preprocessing such as number of bad channels and percentage of bad samples. This denoising step was implemented via a linear model with H as the dependent variable and diagnosis, age, study site, number of bad channels, and percentage of bad samples as independent variables. Beta coefficients for number of bad channels and percentage of bad samples were isolated and then projected out of the data to adjust H values for these metrics. This procedure was done to ensure that any residual effects of data quality were handled before subtyping analysis was implemented. Denoised H data was then utilized in all further downstream analyses.

The primary subtyping analysis was implemented with stability-based relative clustering validation, otherwise known as *reval*[25]. With *reval* we aim to identify unsupervised data-driven

clusters based on H within the autism sample (n = 286) and to identify the optimal number of clusters with cross-validation and test generalizability in new independent data. A 50-50 split was used to separate data into the initial training and validation sets. These sets were balanced for presence or absence of comorbid ADHD diagnosis and age and then preprocessed by scaling it to a mean of 0 and a standard deviation of 1. We also applied a preprocessing step of reducing the dimensionality of the data to 2 dimensions with Uniform Manifold Approximation and Projection (UMAP)[50] (number of neighbors = 30, number of components = 2, minimum distance = 0, random state seed = 24). The UMAP transformations derived from the training set were then applied to the validation set. After preprocessing we implement the primary steps within the reval pipeline - clustering and then producing a classifier that can predict clustering labels on held-out data. The clustering algorithm used throughout is k-means clustering, and the classifier was a k-nearest neighbor classifier (k=3). *reval* utilizes the training data within an internal 2-fold cross-validation loop to identify the optimal k clusters that minimizes clustering stability. This process is repeated 100 times to identify the best cluster solution. Once the optimal k is identified in the internal cross-validation of the training data, a classifier is fit to predict those cluster labels in new data. K-means clustering with the identified optimal k is then applied to the held-out validation set and then the classifier fit to the training data applied to the validation set with the goal of accurately predicting the clustering labels identified in the validation set. Generalization accuracy is then computed by comparing the actual clustering labels of validation set to the classifier's predicted labels. The generalization accuracy computed from *reval* on the validation set informs us as to how robust, stable, and reproducible are clustering labels in the autism population based on EEG H data. However, *reval* itself does not test the hypothesis of whether the clusters come from an underlying distribution a single or multiple Gaussians. The null hypothesis if clusters do not actually exist is that the clustering labels come from a single multivariate Gaussian distribution. Thus, to formally test the data with respect to this null hypothesis, we use the *sigclust* function within the *SigClust* R library to test whether the data deviates from this single multivariate Gaussian null distribution[51]. Finally, given that *reval* clustering is computed twice within the same individuals - once under eyes open and then again for eyes closed condition - this allowed the opportunity to examine consistency of subtype solutions across these two eye conditions. This was achieved descriptively by computing alluvial plots that allow us to visualize the percentage of individuals that stay within the relatively same clustering subtype or which move to a different subtype across the eye conditions. For the individuals who changed subtypes across eyes open and eyes closed, we used linear modeling to test whether these types of individuals differed phenotypically (e.g., on age, FIQ, and autism symptomatology measures) from individuals who stayed within the same subtype.

Once autism EEG-derived H subtypes were identified, we next tested for between-group differences between the autism subtypes and a typically-developing (TD) control group. For these analyses we utilized linear mixed effect models, computed per each electrode with the *lmer* function in the *lme4* R library. The dependent variable in these models were H values. Fixed effects of subtype, age and the subtype*age interaction were modeled as fixed effects of interest. Random effects in the model were acquisition site (random intercepts) and repeated measures such as block (random intercepts and slopes per each subject). Multiple comparisons correction was implemented with false discovery rate (FDR; $q < 0.05$). Because electrode-wise analyses like these may be heavily redundant due to correlations between electrodes, we also utilized principal component analysis (PCA) as a more parsimonious way of summarizing H variation into global versus more localized E:I effects. Examination of scree plots from the PCA analysis showed an

elbow after the first 5 PCs (Supplementary Figure 4), which explained ~85% of the variance in H. Thus, the first 5 PCs were retained for hypothesis testing under the same linear mixed effect model framework as described above. All fixed and random effects were the same as the previously described analysis. The main change is that PCs scores from each of the first 5 PCs are used as the dependent variable in the model. All between-group differences are descriptively shown in figures as standardized effect sizes, operationalized as Cohen's d in standard deviation units.

Behavioral and brain-behavior relationship analyses

Subtypes were tested for behavioral differences on age, FIQ, and autism symptomatology measures (e.g., ASSQ, SCQ, SRS, RBS) using linear models with the behavioral variable as the dependent variable and with subtype as the independent variable. Multiple comparison correction was achieved at FDR $q<0.05$. To test for brain-behavioral relationships we used a multivariate association analysis technique known as partial least squares (PLS) correlation analysis[30], which has been utilized extensively in our past work on autism subtyping[52–54]. For these analyses, we utilized behavioral variables of age, FIQ, ASSQ, SCQ and SRS subscales. The RBS was not utilized in the PLS analysis due to too much missing data. Two PLS analyses were computed - one per eyes open and eyes closed conditions. Groups (TD, autism 1, and autism 2) were input as blocked structures in the PLS analysis. Behavior-PLS analysis was utilized here, whereby H values are the multivariate brain matrix (i.e. the H matrix) and age, FIQ, and autism symptomatology measures are used as the multivariate behavioral matrix. For these analyses, repeat eyes open or eyes closed blocks were concatenated in the H matrix such that each individual or row in the matrix had 93 electrodes * 5 blocks (i.e. 465 H values). Statistical inference on latent variable pairs was achieved in the PLS analysis using permutation analysis (10,000 permutations). Brain bootstrap ratios (BSR) and 95% confidence intervals on behavior scores were computed via bootstrap resampling (10,000 resamples). Latent variable pairs were judged to be statistically significant based on permutation p-values being corrected with FDR $q<0.05$. Behavioral variables with 95% confidence intervals that do not include 0 are considered important for driving the relationship within the identified LV and are thus annotated in the figures describing the PLS results. Brain BSRs are typically considered important when BSRs exceed 2[30].

In-silico modeling of LFP and EEG data

The *in-silico* modeling simulates a recurrently connected network model of leaky-integrate-and-fire (LIF) pyramidal (excitatory) and interneuron (inhibitory) point neurons[16] and computes highly realistic approximations (proxies) of the extracellular potentials (LFP, EEG) and spiking activity[17,55]. This recurrent spiking network model represents a relatively standard model of a recurrent cortical circuit that receives input as recurrent connections between and within populations and receives external inputs. The network structure and parameters are identical to those used in past work[11,17,56] (Supplementary Table 10). The network is composed of 5000 neurons, of which 4000 are excitatory (i.e. they form AMPA-like excitatory synapses with other neurons) and 1000 inhibitory (forming GABA-like synapses). Neurons are randomly connected with a connection probability between each pair of neurons of 0.2. Both populations receive two different types of external Poisson inputs: 1) a constant-rate input giving the network a level of excitation from the outside and 2) a time-fluctuating zero mean input generated by an Ornstein-Uhlenbeck (OU) process, useful to simulate broadband fluctuations in network inputs that may

conceivably arise because of ongoing large-scale brain state changes. The LFP is computed as the sum of absolute values of AMPA and GABA postsynaptic currents on excitatory cells[18,55]. This simple estimation of LFPs was shown to capture more than 90% of variance of both experimental data recorded from cortical field potentials and of simulated LFPs that would have been produced by the same spiking activity within a network of 3D neurons with highly realistic morphology [18,19,55].

In a first set of simulations, we changed across simulations the E:I ratio (g) defined as the ratio between the strengths of the excitatory (g_E) and inhibitory (g_I) synaptic conductances of connections to excitatory neurons. We did so by keeping g_E fixed and changing g_I for both I-to-E and I-to-I connections. This ratio was varied across simulations between 5.7-14 times greater inhibition relative to excitation, a range that reproduces well both spontaneous and stimulus-evoked extracellular potentials and firing regimes in the cortex[11,18,19]. In a second set of simulations, we mimicked the chemogenetic application of hM4Di to all neurons as performed in the DREADD mouse experiments. To achieve this, we simultaneously varied the resting potential (E_l) of both E and I neurons between -70 mV (our reference value) and -75 mV, thus inducing hyperpolarization when making E_l more negative. We also reduced synaptic strength across all connections to emulate synaptic silencing of the hM4Di manipulation[21]. Synaptic strength was reduced proportionally to the decrease of E_l (by 0.1 nS for a decrease of 1 mV in E_l). In a third set of simulations, we mimicked the application of the excitatory DREADD manipulation, CamkII-hM3D(Gq), to the excitatory neurons by varying the resting potential (E_l) of E neurons between -70 mV (our reference value) and -65 mV, thus inducing depolarization when making E_l less negative.

For all simulations the level of constant input (ν0) was set to 2 spikes per second per input cell (800 input cells) in order to resemble resting state conditions, including average firing rates of no more than few spikes/s and an asynchronous irregular firing regime, as observed in cortical spontaneous activity. Using simulated AMPA and GABA currents from the model, we then computed the ERWS2 non-causal EEG proxy as described in prior work[17]. The ERWS2 non-causal EEG proxy used here accounts for approximately 95% of the variance of ground truth EEG signal produced by the same spiking activity within a network of 3D neurons with highly realistic morphology[17]. Using both simulated LFP and EEG proxy data from the model, we computed H identically to how H was computed in human EEG data and the 1/f slope as computed by Gao and colleagues[10]. To verify that increasing g or E_l results in enhanced excitability of the network, we computed Pearson correlations between the parameters (g or E_l) with the mean firing rate of E and I neurons in the model. These same tests were also applied to H, 1/f slope, and total broadband power computed on simulated LFP or EEG. Finally, we examined out-of-sample predictive accuracy for inferring the mean firing rate of the network from H, 1/f slope, or total power computed from EEG and LFP data. To do this, we trained a linear model on either H, 1/f slope, or total power on a 'discovery' dataset of simulated LFP or EEG data. This model was then fitted to H, 1/f slope, or total power computed on independently simulated EEG or LFP dataset (i.e. the 'validation' set) that uses the same simulation parameters, but with resampled connections and noise values. We then compared the ground truth of the firing rate in the validation set to the model's predicted firing rate, measuring strength of association with the R^2 obtained from linear regression fitting.

In-vivo chemogenetic experiments altering excitation-inhibition balance

All in-vivo studies in mice were conducted in accordance with the Italian law (DL 116, 1992 Ministero della Sanità, Roma) and the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. Animal research protocols were also reviewed and consented to by the animal care committee of the Istituto Italiano di Tecnologia and University of Trento.

Experimental procedures for viral expression and chemogenetic manipulations have been described in greater detail elsewhere[22]. Briefly, six-to-eight week old adult male C57Bl6/J mice (Jackson Laboratories; Bar Harbor, ME, USA) were anesthetized with isoflurane (isoflurane 4%) and head-fixed in a mouse stereotaxic apparatus (isoflurane 2%, Stoelting). Viral injections were performed with a Hamilton syringe mounted on Nanoliter Syringe Pump with controller (KD Scientific), at a speed of 0.05 ml/min, followed by a 5–10 min waiting period, to avoid backflow of viral solution and unspecific labeling. Viral suspensions were injected bilaterally in the mouse medial PFC using the following coordinates, expressed in millimeter from bregma: 1.7 AP, ± -0.3 ML, -1.7 DV.

For the DREADD silencing experiment, hM4Di DREADD was transduced using an AAV8-hSyn-hM4D(Gi)-mCherry construct. Control animals were injected with a control AAV8-hSyn-GFP virus (www.addgene.com). These viral suspensions were injected using a 1mL bilateral injection volume in n = 5 hM4Di DREADD and n = 5 SHAM mice, respectively. For the DREADD excitation experiment, CamkII-hM3D(Gq) DREADD was transduced using an AAV8-CamkII-hM3D(Gq)-mCherry construct which we injected bilaterally into the PFC as described above at a volume of 500 µL per hemisphere. Control animals for this experiment underwent a sham surgical procedure during which injection needles were inserted into parenchymal cerebral tissue at the correct stereotaxic coordinates, but no viral suspension was administered. This experimental cohort was composed of n = 5 CamkII-hM3D(Gq) DREADD mice, and n = 8 SHAM. Prior to each DREADD experiment, we waited at least 3 weeks to allow for maximal viral expression.

For hM4Di experiments, electrophysiological recordings were carried out using animal preparation and sedation regimes employed in prior studies[22,57,58]. Mice were anesthetized with isoflurane (4% induction), intubated, artificially ventilated (2% maintenance), and head-fixed in a stereotaxic apparatus (Stoelting). The tail vein was cannulated for clozapine-N-oxide (CNO) injection (2 mg/kg). To ensure maximal consistency between viral injections and recording site, the skull surface was exposed and an insertion hole was gently drilled through the skull corresponding to the location of prior viral injection point. A 16-channel linear probe (Neuronexus, USA) was next inserted through the overlying dura mater by a microdrive array system (Kopf Instruments, Germany) at an insertion rate of 1 µm/min to reach the same stereotaxic coordinates employed for viral injection. Electrode insertion was performed in 3 steps of similar length, with a 20-minute waiting period in between to allow the tissue to reposition after each insertion step. For hM4Di recordings, during the last step of electrode positioning, isoflurane was discontinued and replaced by halothane at a maintenance level of 0.75%. Electrophysiological data acquisition commenced 1 hour after isoflurane cessation. Such transition time was required to ensure complete

washout of isoflurane anesthesia and avoid residual burst-suppressing activity associated with extended exposure to deep anesthetic levels.

CamkII-hM3D(Gq) electrophysiological acquisitions were carried out using a combination of low dose isoflurane and medetomidine[59]. Briefly, animals were anesthetized with isoflurane (4% induction, 2% maintenance), intubated and head fixed on the stereotaxic frame. The tail vein was cannulated for medetomidine infusion. After craniotomy, electrode coordinates and insertion were performed as described before. The tail vein was cannulated for posterior medetomidine infusion, and a second canula was placed intraperitoneally for CNO administration (0.5 mg/kg). Temperature was constantly monitored and maintained to 36.5+-0.5 °C. Craniotomy procedures, and electrode insertion were performed as described before. Following the last step of the electrode insertion, sedation via a combination of medetomidine and isoflurane commenced. An initial intravenous bolus of 0.05 mg/kg of medetomidine was injected, and isoflurane was lowered to 1%. After 5 minutes, isoflurane level was lowered further to 0.3-0.5% and a constant infusion of medetomidine (0.1 mg/kg/h) was applied for maintenance until the end of the acquisition. Neural activity was next recorded in consecutive 5-minute time bins to cover a 15 min pre-injection time window, and a 60 min post CNO timeframe. Signals were amplified using an RHD 2000 amplifier system (Intan Technologies, RHD Recording Controller Software, v2.09) to acquire electrophysiological data at a sampling rate of 20 kHz.

To compute the LFP signal, raw extracellular recordings were first downsampled to 4 kHz, then band-pass filtered to 0.1–250 Hz using a two-step procedure[60]. Briefly, raw timeseries were first low-pass filtered using a 4th order Butterworth filter with a cut-off frequency of 1 kHz. The resulting timeseries were next downsampled to 2 kHz, then again filtered using a Kaiser window filter between 0.1 Hz to 250 Hz (with a sharp transition bandwidth of 1 Hz, passband ripple of 0.01 dB and a stop band attenuation of 60 dB) and then resampled at 1 kHz. Filtering was applied both forward and backward to remove filtering phase transitions lags.

Multi-unity activity (MUA) was computed following the procedure described by Belitski and colleagues[60]. A high-pass filter was applied to the extracellular signal (4th order Butterworth filter with cut off frequency over 100Hz), followed by a band-pass filter between 400 and 3000 Hz using a Kaiser window filter (with transition band of 50 Hz, stopband attenuation of 60 dB, and passband ripple of 0.01 dB). Events over a threshold corresponding to 4-times the median of the signal of all traces (baseline and post injection combined), divided by 0.6745 [61] were counted as neuronal action potentials (spikes). For the final count, spikes were considered to be biologically plausible, and therefore retained, only if occurring more than 1 ms apart.

DREADD LFP data analysis

Preprocessed LFP and MUA data from the DREADD experiments were cut into 4-second segments, with approximately 1 minute of separation between each segment. Given the high level of correlation between the 16 channels of recorded LFP data, we used principal component analysis as a dimensionality reduction technique to capture the vast majority of shared variance (e.g., 95-99%) across channels into one variable (PC1). PC1 was then utilized as input for computing H in a manner identical to how H was computed for human EEG data. The DREADD experiments consisted of 3 phases, called baseline, transition, and treatment. The baseline phase

was a 15-minute time period before the drug injection. The transition phase was defined here as a subsequent 20-minute period post-injection when the drug begins to exert its effect before reaching full pharmacokinetic equilibrations[22,62]. The treatment phase followed the transition phase (duration 25 minutes) and is considered as the time period where the drug is predicted to exert its maximal effect. Before hypothesis test statistical modeling, the mean and standard deviation (SD) of H for each mouse was computed within the baseline period and then used to normalize all H values across the entire experiment as a baseline normalization. This allows the dependent variable in our statistical model to be baseline normalized H, quantifying H in terms of change from baseline average H in units of SD. The same baseline normalization was carried out for the analysis of the 1/f slope, computed between 30-70 Hz, and for the integrated spectral power, defined as the sum of the LFP power spectrum across all frequencies (0-80 Hz).

All statistical modeling for hypothesis testing on LFP H data, 1/f slope, and integrated spectral power for the DREADD experiments was implemented using linear mixed effect models, implemented with the *lmer* function in the *lmerTest* R library. The dependent variable in the model was baseline normalized H. Fixed effects were group (DREADD) and condition (baseline, transition, treatment) and their interaction. Random effects in the model were modeled with random intercepts for each mouse. The same hypothesis testing approach with linear mixed effect models was used for MUA spiking activity data.

Data availability

CMI-HBN data can be found at http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/.

Code availability

The *reval* Python library can be found on GitHub (https://github.com/IIT-LAND/reval_clustering) and the documentation can be found at https://reval.readthedocs.io. Analysis code for the study is available on our GitHub repo (https://github.com/IIT-LAND/CMI_EEG_H). The code for the simulation of point-neuron networks is available in our Gitlab repository (https://gitlab.com/panzerilab/CMI_EEG_networks).

Acknowledgments

This project was supported by funding from the Simons Foundation for Autism Research Initiative (SFARI; grant number 982347) to MVL, AG, and SP and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement number 755816, AUTISMS to MVL; grant agreement number 802371, DISCONN to AG).

Author Contributions

Conceptualization: MVL, AG, SP. Methodology: MVL, NB, AV, VM, AG, SP. Formal analysis: MVL, NB, AV, VM, GML, GM, SBM, SP. Investigation: MVL, NB, DSY, GML, GM, SBM. Writing - original draft preparation: MVL, NB, GML, GM, SBM, DSY, DB, AV, VM, PMC, AG, SP. Writing - review and editing: MVL, NB, GML, GM, SBM, DSY, AV, DB, VM, PMC, AG, SP. Visualization: MVL, NB. Supervision: MVL, AG, SP. Project administration: MVL, AG, SP. Funding acquisition: MVL, AG, SP.

Competing Interests

The authors declare no competing interests.

References

1. Rubenstein, J.L.R., and Merzenich, M.M. (2003). Model of autism: increased ratio of excitation/inhibition in key neural systems. Genes Brain Behav. *2*, 255–267.

2. Nelson, S.B., and Valakh, V. (2015). Excitatory/Inhibitory Balance and Circuit Homeostasis in Autism Spectrum Disorders. Neuron *87*, 684–698.

3. Sohal, V.S., and Rubenstein, J.L.R. (2019). Excitation-inhibition balance as a framework for investigating mechanisms in neuropsychiatric disorders. Mol. Psychiatry *24*, 1248–1257.

4. Willsey, H.R., Willsey, A.J., Wang, B., and State, M.W. (2022). Genomics, convergent neuroscience and progress in understanding autism spectrum disorder. Nat. Rev. Neurosci. *23*, 323–341.

5. Antoine, M.W., Langberg, T., Schnepel, P., and Feldman, D.E. (2019). Increased Excitation-Inhibition Ratio Stabilizes Synapse and Circuit Excitability in Four Autism Mouse Models. Neuron *101*, 648-661.

6. Ahmad, J., Ellis, C., Leech, R., Voytek, B., Garces, P., Jones, E., Buitelaar, J., Loth, E., Dos Santos, F.P., Amil, A.F., et al. (2022). From mechanisms to markers: novel noninvasive EEG proxy markers of the neural excitation and inhibition system in humans. Transl. Psychiatry *12*, 467.

7. Shafiei, G., Fulcher, B.D., Voytek, B., Satterthwaite, T.D., Baillet, S., and Misic, B. (2023). Neurophysiological signatures of cortical micro-architecture. Nat. Commun. *14*, 6000.

8. Webb, S.J., Naples, A.J., Levin, A.R., Hellemann, G., Borland, H., Benton, J., Carlos, C., McAllister, T., Santhosh, M., Seow, H., et al. (2023). The Autism Biomarkers Consortium for Clinical Trials: Initial Evaluation of a Battery of Candidate EEG Biomarkers. Am. J. Psychiatry *180*, 41–49.

9. Mason, L., Moessnang, C., Chatham, C., Ham, L., Tillmann, J., Dumas, G., Ellis, C., Leblond, C.S., Cliquet, F., Bourgeron, T., et al. (2022). Stratifying the autistic phenotype using electrophysiological indices of social perception. Sci. Transl. Med. *14*, eabf8987.

10. Gao, R., Peterson, E.J., and Voytek, B. (2017). Inferring synaptic excitation/inhibition balance from field potentials. Neuroimage *158*, 70–78.

11. Trakoshis, S., Martínez-Cañada, P., Rocchi, F., Canella, C., You, W., Chakrabarti, B., Ruigrok, A.N., Bullmore, E.T., Suckling, J., Markicevic, M., et al. (2020). Intrinsic excitation-inhibition imbalance affects medial prefrontal cortex differently in autistic men versus women. eLife *9*, e55684.

12. Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K.D. (2010). The asynchronous state in cortical circuits. Science *327*, 587–590.

13. Ahmadian, Y., and Miller, K.D. (2021). What is the dynamical regime of cerebral cortex? Neuron *109*, 3373–3391.

14. Manning, J.R., Jacobs, J., Fried, I., and Kahana, M.J. (2009). Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. J. Neurosci. *29*, 13613–13620.

15. Hong, S.-J., Vogelstein, J.T., Gozzi, A., Bernhardt, B.C., Yeo, B.T.T., Milham, M.P., and Di Martino, A. (2020). Toward Neurosubtypes in Autism. Biol. Psychiatry *88*, 111–128.

16. Brunel, N., and Wang, X.-J. (2003). What determines the frequency of fast network oscillations with irregular neural discharges? I. Synaptic dynamics and excitation-inhibition balance. J. Neurophysiol. *90*, 415–430.

17. Martínez-Cañada, P., Ness, T.V., Einevoll, G.T., Fellin, T., and Panzeri, S. (2021). Computation of the electroencephalogram (EEG) from network models of point neurons. PLoS Comput. Biol. *17*, e1008893.

18. Mazzoni, A., Panzeri, S., Logothetis, N.K., and Brunel, N. (2008). Encoding of naturalistic stimuli by local field potential spectra in networks of excitatory and inhibitory neurons. PLoS Comput. Biol. *4*, e1000239.

19. Barbieri, F., Mazzoni, A., Logothetis, N.K., Panzeri, S., and Brunel, N. (2014). Stimulus dependence of local field potential spectra: experiment versus theory. J. Neurosci. *34*, 14589–14605.

20. Einevoll, G.T., Kayser, C., Logothetis, N.K., and Panzeri, S. (2013). Modelling and analysis of local field potentials for studying the function of cortical circuits. Nat. Rev. Neurosci. *14*, 770–785.

21. Roth, B.L. (2016). DREADDs for Neuroscientists. Neuron *89*, 683–694.

22. Rocchi, F., Canella, C., Noei, S., Gutierrez-Barragan, D., Coletta, L., Galbusera, A., Stuefer, A., Vassanelli, S., Pasqualetti, M., Iurilli, G., et al. (2022). Increased fMRI connectivity upon chemogenetic inhibition of the mouse prefrontal cortex. Nat. Commun. *13*, 1056.

23. Logothetis, N.K. (2008). What we can do and what we cannot do with fMRI. Nature *453*, 869–878.

24. Buzsáki, G. (2006). Rhythms of the brain (Oxford University Press).

25. Landi, I., Mandelli, V., and Lombardo, M.V. (2021). reval: A Python package to determine best clustering solutions with stability-based relative clustering validation. Patterns *2*, 100228.

26. Alexander, L.M., Escalera, J., Ai, L., Andreotti, C., Febre, K., Mangone, A., Vega-Potler, N., Langer, N., Alexander, A., Kovacs, M., et al. (2017). An open resource for transdiagnostic research in pediatric mental health and learning disorders. Sci. Data *4*, 170181.

27. Voytek, B., Kramer, M.A., Case, J., Lepage, K.Q., Tempesta, Z.R., Knight, R.T., and Gazzaley, A. (2015). Age-Related Changes in 1/f Neural Electrophysiological Noise. J. Neurosci. *35*, 13257–13265.

28. Donoghue, T., Haller, M., Peterson, E.J., Varma, P., Sebastian, P., Gao, R., Noto, T., Lara, A.H., Wallis, J.D., Knight, R.T., et al. (2020). Parameterizing neural power spectra into periodic and aperiodic components. Nat. Neurosci. *23*, 1655–1665.

29. Schaworonkow, N., and Voytek, B. (2021). Longitudinal changes in aperiodic and periodic activity in electrophysiological recordings in the first seven months of life. Dev. Cogn. Neurosci. *47*, 100895.

30. Krishnan, A., Williams, L.J., McIntosh, A.R., and Abdi, H. (2011). Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. Neuroimage *56*, 455–475.

31. Colombo, M.A., Napolitani, M., Boly, M., Gosseries, O., Casarotto, S., Rosanova, M., Brichant, J.-F., Boveroux, P., Rex, S., Laureys, S., et al. (2019). The spectral exponent of the resting EEG indexes the presence of consciousness during unresponsiveness induced by propofol, xenon, and ketamine. Neuroimage *189*, 631–644.

32. Mandelli, V., Landi, I., Busuoli, E.M., Courchesne, E., Pierce, K., and Lombardo, M.V. (2023). Prognostic early snapshot stratification of autism based on adaptive functioning. Nat. Ment. Health *1*, 327–336.

33. Mandelli, V., Landi, I., Ceccarelli, S.B., Molteni, M., Nobile, M., D'Ausilio, A., Fadiga, L., Crippa, A., and Lombardo, M.V. (2023). Enhanced motor noise in an autism subtype with poor motor skills. medRxiv. https://doi.org/10.1101/2023.03.25.23287738.

34. Peck, F.C., Gabard-Durnam, L.J., Wilkinson, C.L., Bosl, W., Tager-Flusberg, H., and Nelson, C.A. (2021). Prediction of autism spectrum disorder diagnosis using nonlinear measures of language-related EEG at 6 and 12 months. J. Neurodev. Disord. *13*, 57.

35. Catarino, A., Churches, O., Baron-Cohen, S., Andrade, A., and Ring, H. (2011). Atypical EEG complexity in autism spectrum conditions: a multiscale entropy analysis. Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol. *122*, 2375–2383. https://doi.org/10.1016/j.clinph.2011.05.004.

36. Lai, M.-C., Lombardo, M.V., Chakrabarti, B., Sadek, S.A., Pasco, G., Wheelwright, S.J., Bullmore, E.T., Baron-Cohen, S., MRC AIMS Consortium, and Suckling, J. (2010). A shift to randomness of brain oscillations in people with autism. Biol. Psychiatry *68*, 1092–1099. https://doi.org/10.1016/j.biopsych.2010.06.027.

37. Uscătescu, L.C., Hyatt, C.J., Dunn, J., Kronbichler, M., Calhoun, V., Corbera, S., Pelphrey, K., Pittman, B., Pearlson, G., and Assaf, M. (2022). Using the Excitation/Inhibition Ratio to Optimize the Classification of Autism and Schizophrenia. medRxiv. https://doi.org/10.1101/2022.05.24.22275531.

38. Watanabe, T., Rees, G., and Masuda, N. (2019). Atypical intrinsic neural timescale in autism. eLife *8*, e42256.

39. Sydnor, V.J., Larsen, B., Bassett, D.S., Alexander-Bloch, A., Fair, D.A., Liston, C., Mackey, A.P., Milham, M.P., Pines, A., Roalf, D.R., et al. (2021). Neurodevelopment of the association cortices: Patterns, mechanisms, and implications for psychopathology. Neuron *109*, 2820–2846.

40. Gao, R., van den Brink, R.L., Pfeffer, T., and Voytek, B. (2020). Neuronal timescales are functionally dynamic and shaped by cortical microarchitecture. eLife *9*, e61277.

41. Wang, X.-J. (2020). Macroscopic gradients of synaptic excitation and inhibition in the neocortex. Nat. Rev. Neurosci. *21*, 169–178.

42. Constantino, J.N., and Todd, R.D. (2003). Autistic traits in the general population: a twin study. Arch. Gen. Psychiatry *60*, 524–530.

43. Antaki, D., Guevara, J., Maihofer, A.X., Klein, M., Gujral, M., Grove, J., Carey, C.E., Hong, O., Arranz, M.J., Hervas, A., et al. (2022). A phenotypic spectrum of autism is attributable to the combined effects of rare variants, polygenic risk and sex. Nat. Genet. *54*, 1284–1292.

44. Robertson, C.E., Ratai, E.-M., and Kanwisher, N. (2016). Reduced GABAergic Action in the Autistic Brain. Curr. Biol. *26*, 80–85.

45. Spiegel, A., Mentch, J., Haskins, A.J., and Robertson, C.E. (2019). Slower Binocular Rivalry in the Autistic Brain. Curr. Biol. *29*, 2948-2953.

46. Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods *134*, 9–21.

47. Kothe, C.A.E., and Jung, T.-P. (2016). Artifact removal techniques with signal reconstruction.

48. Chang, C.-Y., Hsu, S.-H., Pion-Tonachini, L., and Jung, T.-P. (2020). Evaluation of Artifact Subspace Reconstruction for Automatic Artifact Components Removal in Multi-Channel EEG Recordings. IEEE Trans. Biomed. Eng. *67*, 1114–1121.

49. Pion-Tonachini, L., Kreutz-Delgado, K., and Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. Neuroimage *198*, 181–197.

50. McInnes, L., Healy, J., and Melville, J. (2020). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv.

51. Liu, Y., Hayes, D.N., Nobel, A., and Marron, J.S. (2008). Statistical Significance of Clustering for High-Dimension, Low–Sample Size Data. J. Am. Stat. Assoc. *103*, 1281–1293.

52. Lombardo, M.V., Pierce, K., Eyler, L.T., Carter Barnes, C., Ahrens-Barbeau, C., Solso, S., Campbell, K., and Courchesne, E. (2015). Different functional neural substrates for good and poor language outcome in autism. Neuron *86*, 567–577.

53. Lombardo, M.V., Pramparo, T., Gazestani, V., Warrier, V., Bethlehem, R.A.I., Carter Barnes, C., Lopez, L., Lewis, N.E., Eyler, L., Pierce, K., et al. (2018). Large-scale associations between the leukocyte transcriptome and BOLD responses to speech differ in autism early language outcome subtypes. Nat. Neurosci. *21*, 1680–1688.

54. Lombardo, M.V., Eyler, L., Pramparo, T., Gazestani, V.H., Hagler, D.J., Chen, C.-H., Dale, A.M., Seidlitz, J., Bethlehem, R.A.I., Bertelsen, N., et al. (2021). Atypical genomic cortical patterning in autism with poor early language outcome. Sci. Adv. *7*, eabh1663.

55. Mazzoni, A., Lindén, H., Cuntz, H., Lansner, A., Panzeri, S., and Einevoll, G.T. (2015). Computing the Local Field Potential (LFP) from Integrate-and-Fire Network Models. PLoS Comput. Biol. *11*, e1004584.

56. Cavallari, S., Panzeri, S., and Mazzoni, A. (2014). Comparison of the dynamics of neural interactions between current-based and conductance-based integrate-and-fire recurrent networks. Front. Neural Circuits *8*, 12.

57. Ferrari, L., Turrini, G., Crestan, V., Bertani, S., Cristofori, P., Bifone, A., and Gozzi, A. (2012). A robust experimental protocol for pharmacological fMRI in rats and mice. J. Neurosci. Methods *204*, 9–18.

58. Gutierrez-Barragan, D., Singh, N.A., Alvino, F.G., Coletta, L., Rocchi, F., De Guzman, E., Galbusera, A., Uboldi, M., Panzeri, S., and Gozzi, A. (2022). Unique spatiotemporal fMRI dynamics in the awake mouse brain. Curr. Biol. *32*, 631-644.

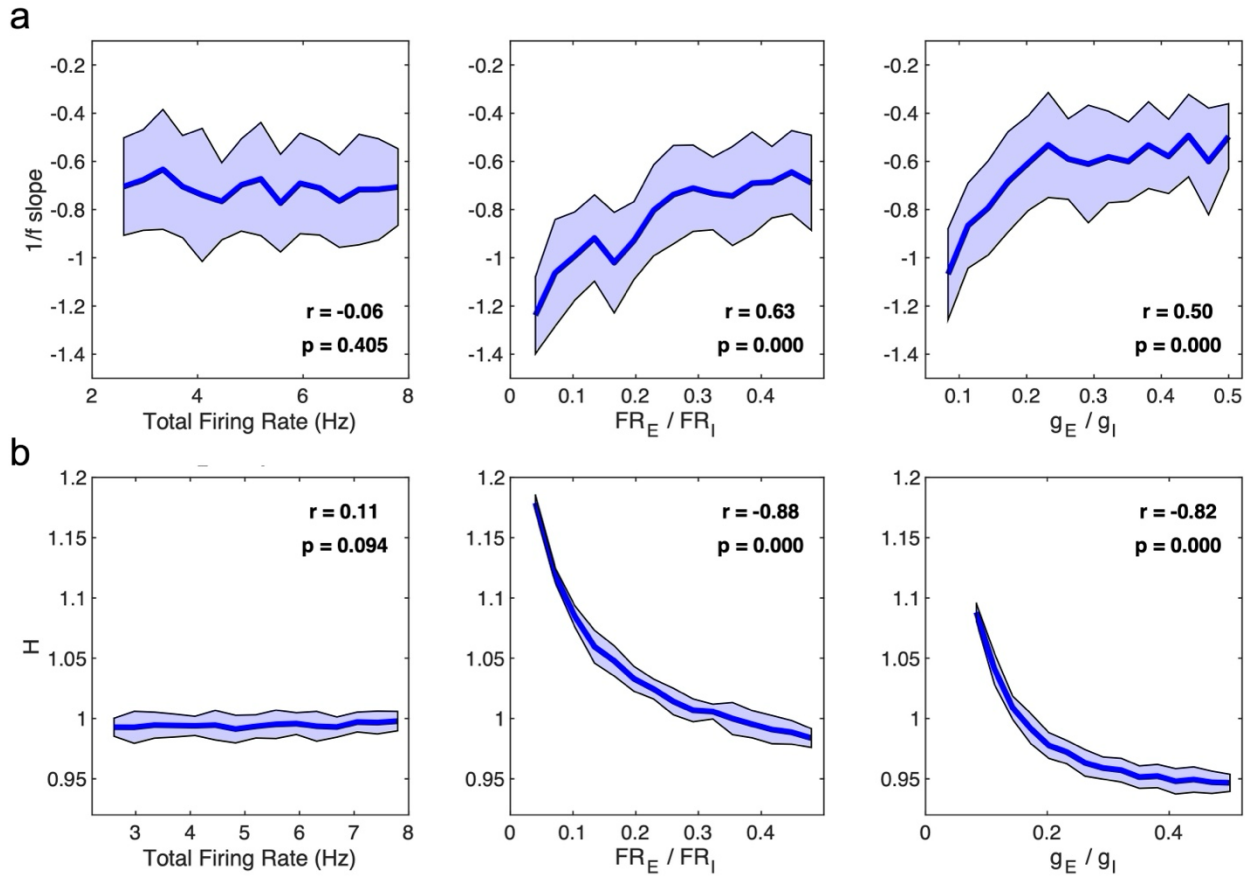
59. Grandjean, J., Canella, C., Anckaerts, C., Ayrancı, G., Bougacha, S., Bienert, T., Buehlmann, D., Coletta, L., Gallino, D., Gass, N., et al. (2020). Common functional networks in the mouse brain revealed by multi-centre resting-state fMRI analysis. Neuroimage *205*, 116278.

60. Belitski, A., Gretton, A., Magri, C., Murayama, Y., Montemurro, M.A., Logothetis, N.K., and Panzeri, S. (2008). Low-frequency local field potentials and spikes in primary visual cortex convey independent visual information. J. Neurosci. *28*, 5696–5709.

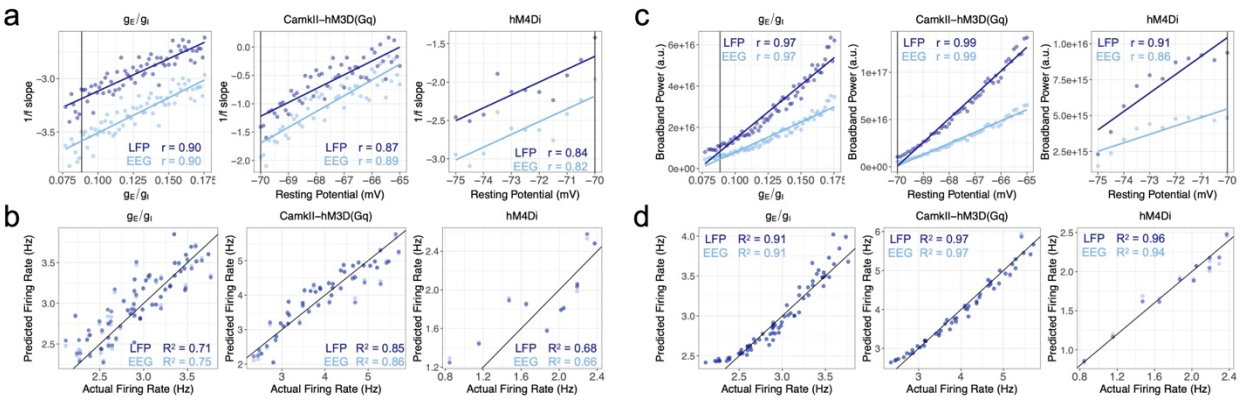
61. Quiroga, R.Q., Nadasdy, Z., and Ben-Shaul, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. Neural Comput. *16*, 1661–1687.

62. Gomez, J.L., Bonaventura, J., Lesniak, W., Mathews, W.B., Sysa-Shah, P., Rodriguez, L.A., Ellis, R.J., Richie, C.T., Harvey, B.K., Dannals, R.F., et al. (2017). Chemogenetics revealed: DREADD occupancy and activation via converted clozapine. Science *357*, 503–507.

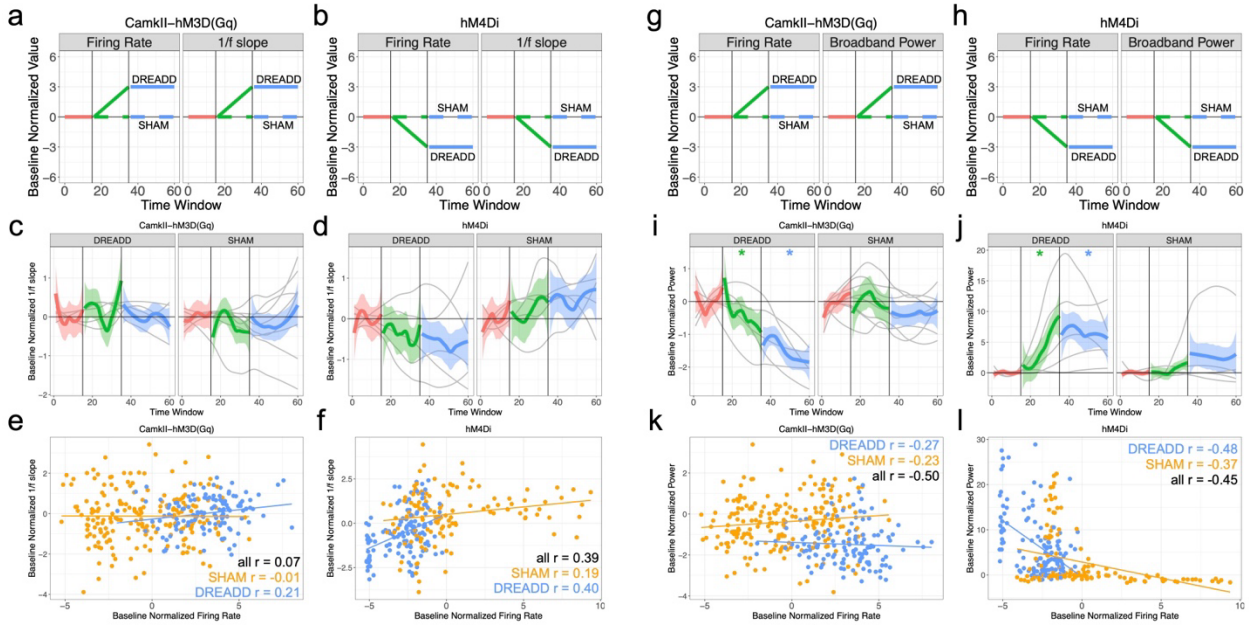
Supplementary Figures



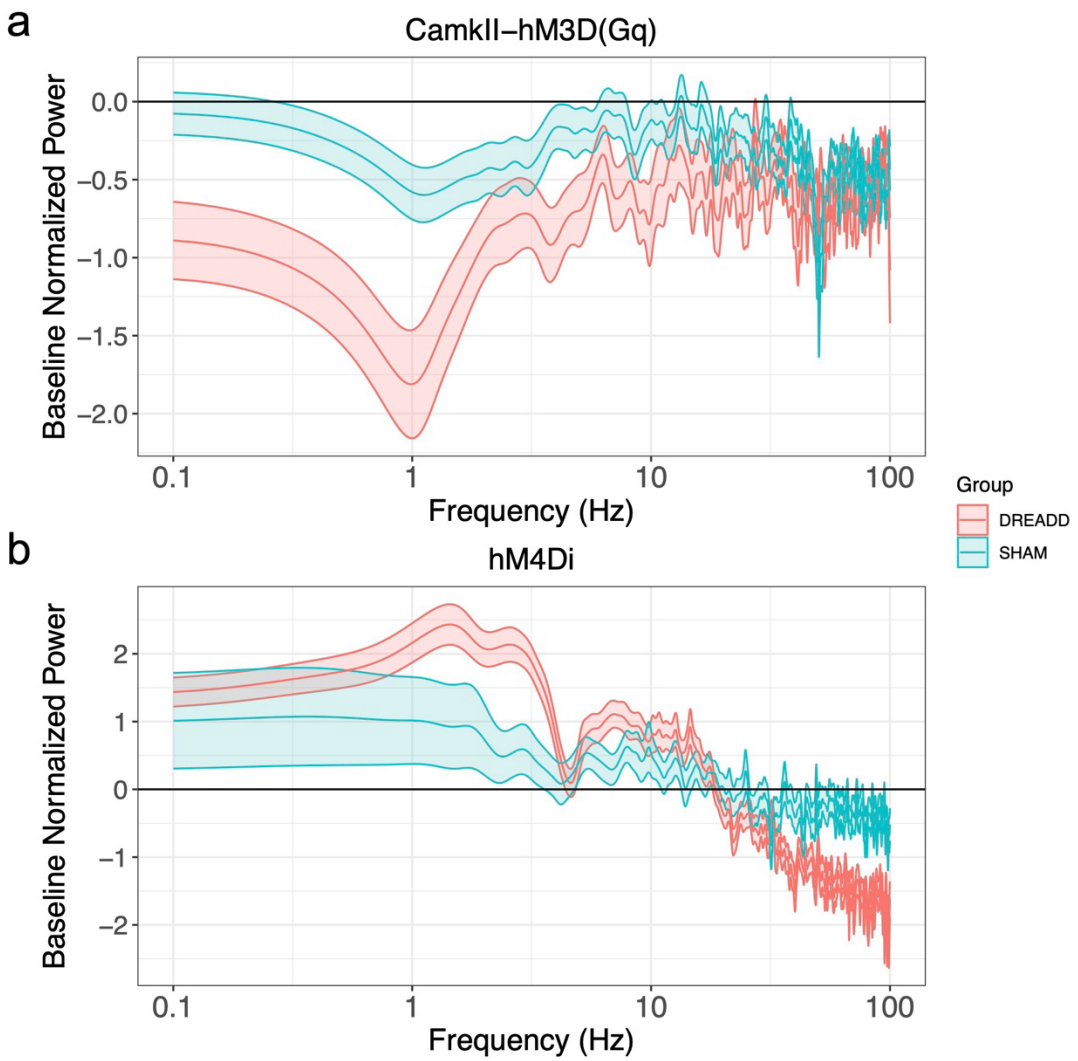
Supplementary Figure 1: Analysis of E:I LFP proxies from the uncoupled model from Gao et al[10]. *This figure presents the results of the analysis of how the LFP proxies of E:I imbalance depend on neural parameters in the uncoupled excitation-inhibition model described by Gao et al[10]. In this model, E and I spikes are generated at random times (homogeneous Poisson process) with a given mean firing rate of each cell type (FR_E for E neurons; FR_I for I neurons). Post-synaptic responses are created exactly as in Gao et al[10], by convolving the spike trains with a beta function. The beta function has parameters τ rise= 0.1ms and 0.5 ms, and τ decay = 2ms and 10 ms for AMPA and GABA synapses respectively. The beta function is multiplied by a synaptic efficacy parameter g_E for E neurons and g_I for I neurons, and the LFP is computed as a sum of all synaptic currents, exactly as in Gao et al[10]. Panel A shows dependence of the 1/f slope (x-axis), computed as linear interpolation as in Gao et al[10] from the LFP in the 30-70 Hz range, on the grand average firing rate defined as the ratio of the firing of all E and I neurons (left), on the ratio between E and I firing rate (center), and on the ratio of synaptic efficacies g_E/g_I (right). Panel B shows the same plots as panel A, but for H (x-axis) rather than for the 1/f slope. In all panels, the line plots the mean and the shaded are the SD over n=15 simulations (10 seconds of simulated activity per simulation) for each parameter value. Each plot shows the Pearson's correlation and corresponding p-value for the association between plotted variables. Both H and 1/f slope in this model correlate tightly with variations in E:I conductance or firing rate ratio, but not with the total firing rate. In contrast, in both our in-silico coupled E:I network model and in-vivo mouse data, H and the 1/f slope correlate with the total firing rate.*



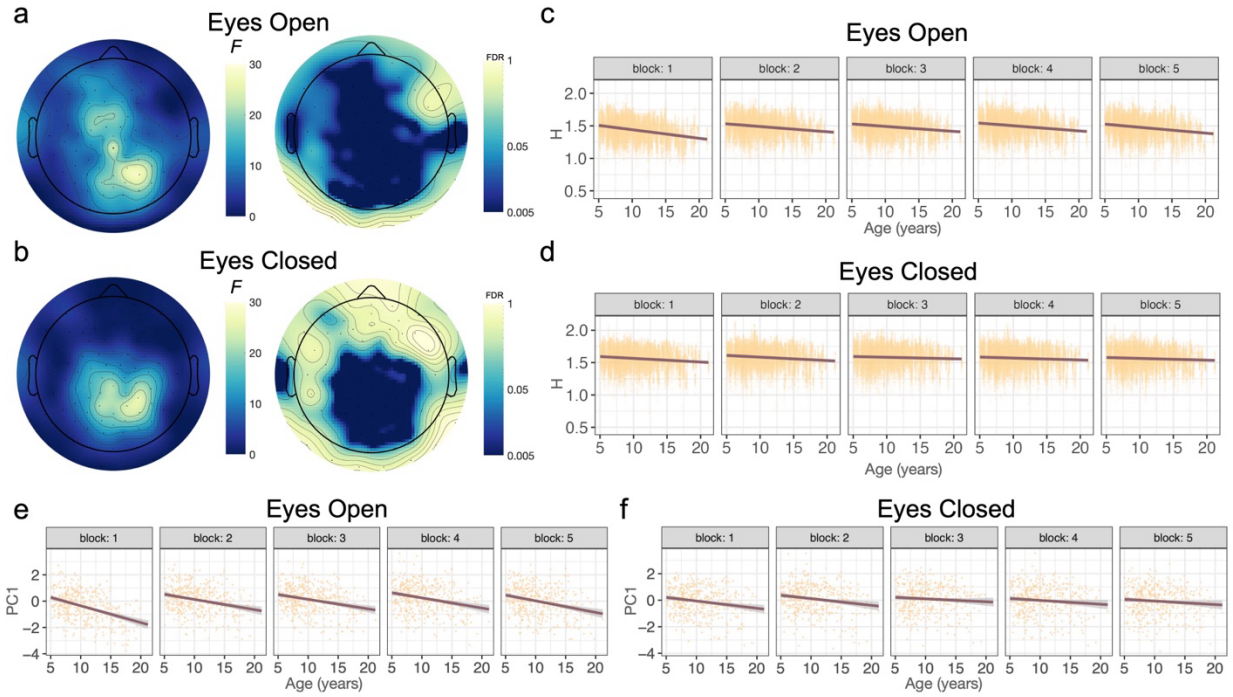
Supplementary Figure 2: *This figure shows how metrics such as 1/f slope and total broadband spectrum power relate to E:I ratio (g) and manipulations of the resting potential (El) in the CamkII-hM3D(Gq) and hM4Di simulations. Panels A and C show scatterplots and Pearson's correlation values (full set of statistics can be found in Supplementary Table 1). Panels B and D plot actual versus out-of-sample predicted firing rates.*



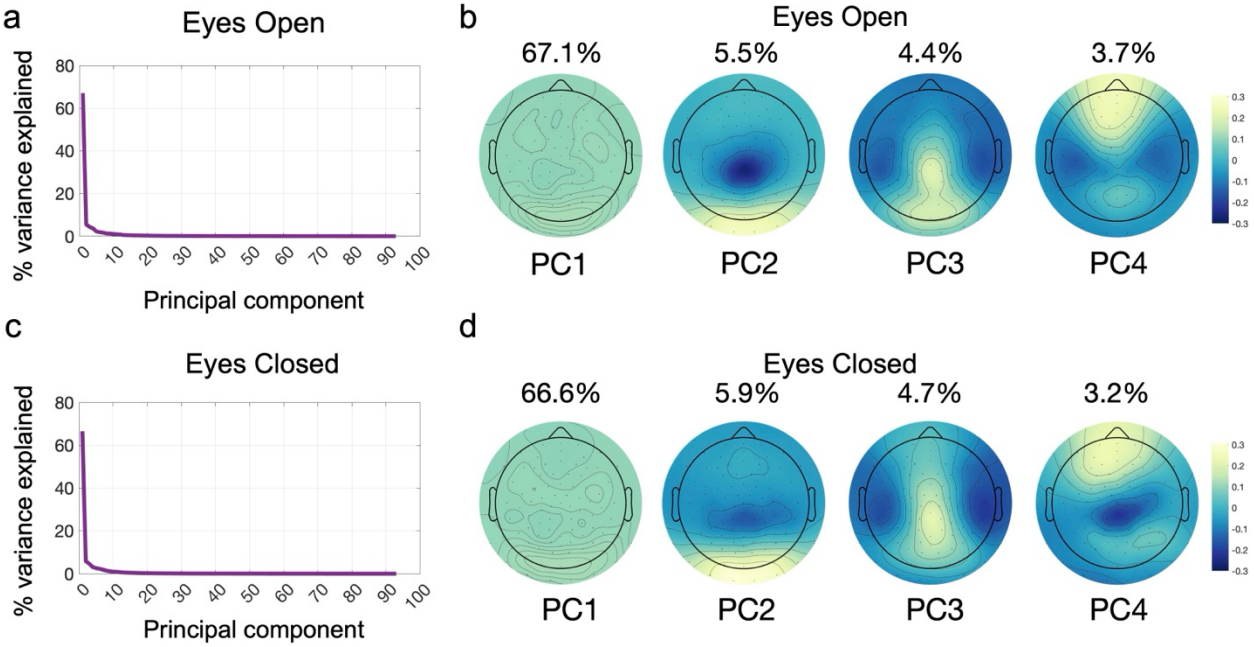
Supplementary Figure 3: *This figure shows how metrics such as 1/f slope and total broadband power change as a function of the two DREADD manipulations that either increase (CamkII-hM3D(Gq)) or decrease (hM4Di) excitability. Panels A-F show results for 1/f slope, while panels G-L show results for total broadband power. Panels A-B and G-H show theoretical predictions informed by the in-silico modeling. Panels C-D and I-J show plots of the metrics as they evolve over time across the 3 phases of the experiment (baseline, transition, treatment). Panels E-F and K-L show scatterplots of firing rate (x-axis) versus the metric (y-axis) when extracted exclusively from the Treatment phase of the experiment.*



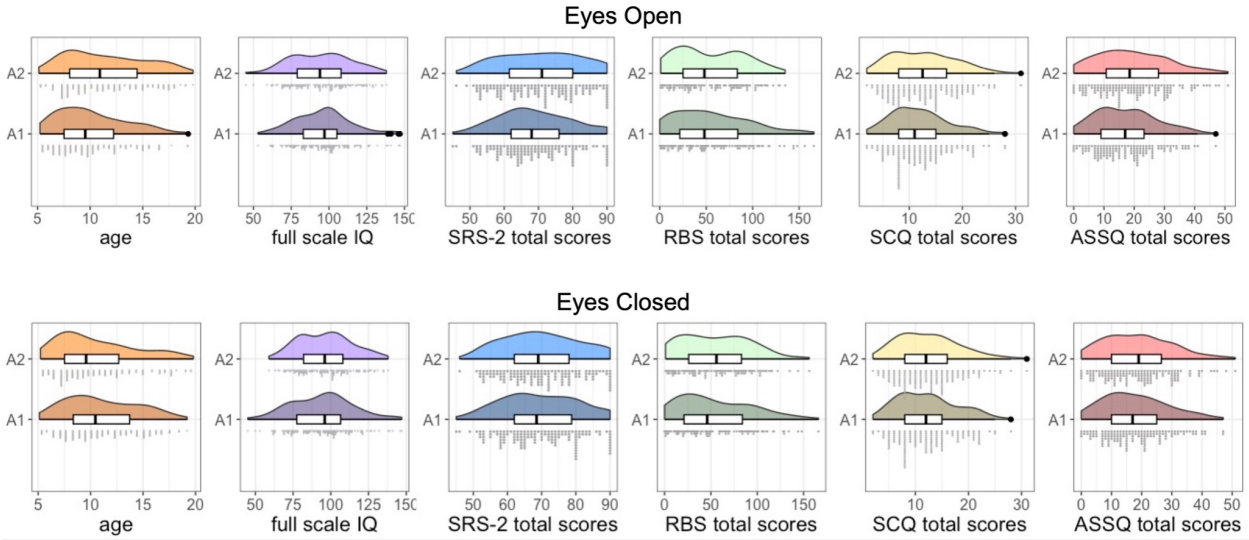
Supplementary Figure 4: Modulations of power spectra of mouse PFC LFPs during the treatment phase of DREADD experiments. *The line and shaded show the mean and SEM across mice and analyzed time windows of baseline-normalized power spectra of the mouse PFC LFPs during the "treatment" phase of DREADD excitation (CamkII-hM3D(Gq); panel A) and silencing (hM4Di; panel B) experiments respectively. Baseline-normalized spectra are obtained by Z-scoring the log of the power spectrum at each frequency measured during the treatment phase with the mean and SD of the log of the power spectra at the corresponding frequency computed during the baseline period. For both types of chemogenetic drugs, we show the spectral modulations both for the saline (SHAM) and CNO (DREADD) manipulations. For application of hM4Di (panel B), we obtain a decrease of higher-frequency (30-70 Hz) oscillations (capturing the reduced local activity and excitability) and an increase of low-frequency 0.1-5 Hz oscillations (capturing the higher locking of local activity to global low-frequency oscillations[22]), compared to SHAM. For application of CamkII-hM3D(Gq) (panel A), we obtain a decrease of low-frequency 0.1-5 Hz oscillations (capturing the weaker locking of local activity to global low-frequency oscillations), compared to SHAM. The changes in low-frequency power induced by the manipulation are opposite in sign with respect to the firing rate changes induced by the same manipulation. This prevents using the total broadband power as a reliable biomarker of the underlying firing rates.*



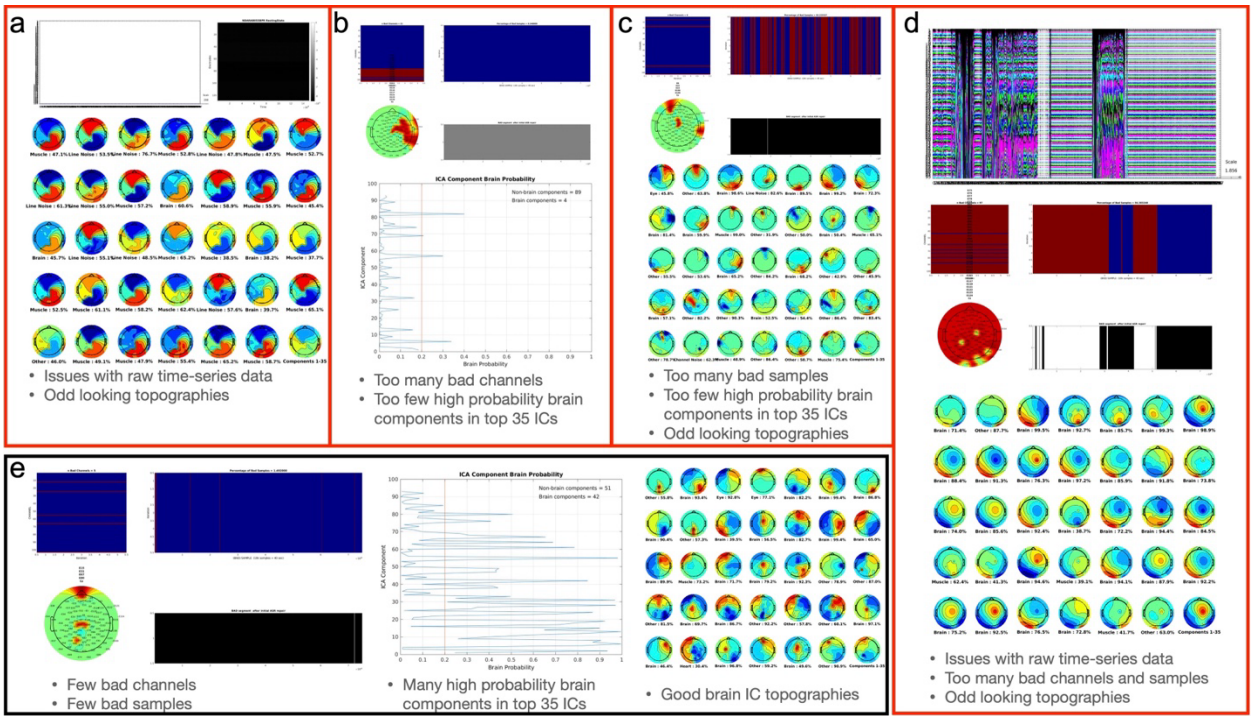
Supplementary Figure 5: Decreasing H with increasing age. *Panels A and B show topoplots of F-statistics of the main effect of age (left) along with FDR topoplots (right) to indicate which electrodes show such significant effects. Panels C and D show scatterplots of H (y-axis) over age (x-axis) for each block in eyes open (C) and eyes closed (D) conditions. Panels E and F show scatterplots of age (x-axis) by PC1 scores (y-axis) for eyes open (E) and eyes closed (F) conditions.*



Supplementary Figure 6: PCA analysis isolates top 4 PCs explaining ~80% of the variance in scalp-wide H. *Panels A and C show scree plots of percentage variance explained (y-axis) for each PC (x-axis) for eyes open (A) and eyes closed (C) conditions. Panels B and D show topoplots of the loadings for eyes open (B) and eyes closed (C) conditions.*



Supplementary Figure 7: Raincloud plots of behavioral variables for each autism E:I neurosubtype. *This plot shows the distribution of values over age, FIQ, and autism symptomatology measures (SRS, RBS, SCQ, and ASSQ) for eyes open (top) and eyes closed (bottom) conditions for both autism subtypes 1 (A1) and 2 (A2).*



Supplementary Figure 8: Examples of issues flagged up in data quality control analysis that warranted exclusion (A-D) or inclusion of data into further downstream analysis. *Panels A-D show different examples of issues identified in data quality inspection at different step of the preprocessing pipeline. Panel E shows an example of a participant with relatively clean data and which was included in further downstream analysis.*

Supplementary Table Legends

Supplementary Table 1*: This table provides all of the correlations (Pearson's r) and p-values from statistical analyses depicted in Figure 1E-G and Supplementary Figure 2.*

Supplementary Table 2*: This table provides all of the statistics from linear models (F-statistics and p-values) run on DREADD experiment data (CamkII-hM3D(Gq) and hM4Di) depicted in Figure 2D-E, 2G-H and Supplementary Figure 3C-D, 3I-J.*

Supplementary Table 3*: This table provides all of the correlations (Pearson's r) and p-values from analyses of trial-level data from the treatment phase of the DREADD experiments depicted in Figure 2F, 2I and Supplementary Figure 3E, 3F, 3K, and 3L.*

Supplementary Table 4*: This table provides p-values from the SigClust analysis for eyes open and closed (rows) and training and validation sets (columns).*

Supplementary Table 5*: This table provides statistics comparing individuals within autism E:I neurosubtypes that are consistent or inconsistent across resting state eyes open versus closed conditions.*

Supplementary Table 6*: This table provides statistics per each electrode for main effects of subtype, age, and the subtype*age interaction, and also pair-wise autism subtype vs TD comparisons.*

Supplementary Table 7*: This table provides statistics per each principal component for main effects of subtype, age, and the subtype*age interaction, pair-wise autism subtype vs TD comparisons, and effect sizes per each block of eyes open and closed conditions.*

Supplementary Table 8*: This table provides descriptive statistics per each demographic or behavioral variable are shown in the table at the top. Directly below the descriptive statistics table are inferential statistics for a model testing for a between-group difference between TD, A1, and A2 (e.g., F-statistics, p-values, FDR). Below this table are further inferential statistical results for follow-up pair-wise comparisons between all groups are reported (e.g., t-statistics, p-values, FDR, Cohen's d). analyzing main effects of subtype, age, and the subtype*age interaction, pair-wise autism subtype vs TD comparisons, and effect sizes per each block of eyes open and closed conditions.*

Supplementary Table 9*: This table shows sample size (n) or mean and standard deviation (in parentheses) for autism E:I neurosubtypes and TD groups. Abbreviations: FIQ: full-scale intelligence quotient, ASSQ: Autism Spectrum Screening Questionnaire; RBS-R: Repetitive Behaviors Scale - Revised (RBS-R); SCQ: Social Communication Questionnaire; SRS-2: Social Responsiveness Scale 2.*

Supplementary Table 10: *This table provides baseline reference parameters of the in-silico recurrent network model.*