Title: Intrinsic excitation-inhibition imbalance affects medial prefrontal cortex differently in autistic men versus women

Authors: Stavros Trakoshis[1,2]*, Pablo Martínez-Cañada[3,4]*, Federico Rocchi[5], Carola Canella[5], Wonsang You[6], Bhismadev Chakrabarti[7,8], Amber N. V. Ruigrok[7], Edward T. Bullmore[9,10], John Suckling[10], Marija Markicevic[11], Valerio Zerbi[11,12], MRC AIMS Consortium†, Simon Baron-Cohen[7,9], Alessandro Gozzi[5]**, Meng-Chuan Lai[7,13,14,15,16]**, Stefano Panzeri[3]**, & Michael V. Lombardo[1,7]**

Affiliations:

1 Laboratory for Autism and Neurodevelopmental Disorders, Center for Neuroscience and Cognitive Systems @UniTn, Istituto Italiano di Tecnologia, Corso Bettini, 31, 38068 Rovereto, Italy
2 Department of Psychology, University of Cyprus, Nicosia, Cyprus
3 Neural Computation Laboratory, Center for Neuroscience and Cognitive Systems @UniTn, Istituto Italiano di Tecnologia, Corso Bettini, 31, 38068 Rovereto, Italy
4 Optical Approaches to Brain Function Laboratory, Department of Neuroscience and Brain Technologies, Istituto Italiano di Tecnologia, Via Morego 30, 16163 Genova, Italy
5 Functional Neuroimaging Laboratory, Center for Neuroscience and Cognitive Systems @UniTn, Istituto Italiano di Tecnologia, Corso Bettini 31, 38068 Rovereto, Italy
6 Children's National Medical Center, Department of Diagnostic Imaging and Radiology, Washington, DC, USA
7 Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom
8 Centre for Integrative Neuroscience and Neurodynamics, School of Psychology and Clinical Language Sciences, University of Reading, Reading, United Kingdom
9 Cambridgeshire and Peterborough National Health Service Foundation Trust, Cambridge, United Kingdom
10 Brain Mapping Unit, Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom
11 Neural Control of Movement Lab, D-HEST, ETH Zurich, Auguste-Piccard-Hof 1, 8093 Zurich, Switzerland
12 Neuroscience Center Zurich, University and ETH Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
13 The Margaret and Wallace McCain Centre for Child, Youth & Family Mental Health, Azrieli Adult Neurodevelopmental Centre, and Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Canada
14 Centre for Brain and Mental Health and Department of Psychiatry, The Hospital for Sick Children, Toronto, Canada
15 Department of Psychiatry, Faculty of Medicine, University of Toronto, Toronto, Canada
16 Department of Psychiatry, National Taiwan University Hospital and College of Medicine, Taipei, Taiwan

* Equal first author contributions
** Equal senior author contributions

Corresponding Author: Michael V. Lombardo (michael.lombardo@iit.it)

† MRC AIMS Consortium Author List (Alphabetical Order)

Anthony J. Bailey (Oxford), Simon Baron-Cohen (Cambridge), Patrick F Bolton (IoPPN), Edward T. Bullmore (Cambridge), Sarah Carrington (Oxford), Marco Catani (IoPPN), Bhismadev Chakrabarti (Cambridge), Michael C. Craig (IoPPN), Eileen M. Daly (IoPPN), Sean C. L. Deoni (IoPPN), Christine Ecker (IoPPN), Francesca Happé (IoPPN), Julian Henty (Cambridge), Peter Jezzard (Oxford), Patrick Johnston (IoPPN), Derek K. Jones (IoPPN), Meng-Chuan Lai (Cambridge), Michael V. Lombardo (Cambridge), Anya Madden (IoPPN), Diane Mullins (IoPPN), Clodagh M. Murphy (IoPPN), Declan G. M. Murphy (IoPPN), Greg Pasco (Cambridge), Amber N. V. Ruigrok (Cambridge), Susan A. Sadek (Cambridge), Debbie Spain (IoPPN), Rose Stewart (Oxford), John Suckling (Cambridge), Sally J. Wheelwright (Cambridge) and Steven C. Williams (IoPPN).

Abstract

Excitation-inhibition (E:I) imbalance is theorized as an important pathophysiological mechanism in autism. Autism affects males more than females and sex-related mechanisms (e.g., X-linked genes, androgen hormones) can influence E:I balance. This suggests that E:I imbalance may affect autism differently in males versus females. With a combination of *in-silico* modeling and *in-vivo* chemogenetic manipulations in mice, we first show that a time-series metric estimated from fMRI BOLD signal, the Hurst exponent (H), can be an index for underlying change in the synaptic E:I ratio. In autism we find that H is reduced, indicating increased excitation, in the medial prefrontal cortex (MPFC) of autistic males but not females. Increasingly intact MPFC H is also associated with heightened ability to behaviorally camouflage social-communicative difficulties, but only in autistic females. This work suggests that H in BOLD can index synaptic E:I ratio and that E:I imbalance affects autistic males and females differently.

Keywords: autism; heterogeneity; excitation; inhibition; fMRI; Hurst exponent; DREADD; sex/gender; camouflaging

Excitation-inhibition (E:I) balance in the brain has been hypothesized to be atypical in many neuropsychiatric conditions[1,2], including autism. Rubenstein and Merzenich originally suggested that some types of autism may be explained by an E:I imbalance that may lead to hyper-excitability in cortical circuitry and potentially enhanced levels of neuronal noise[1]. However, coming to a better understanding of how E:I balance is affected across a heterogeneous mixture of autistic individuals has proven to be challenging because of the limited availability of robust E:I biomarkers that are non-invasive and applicable in humans and which can be measured on a large scale. A majority of the literature about E:I balance in autism extends from investigations of prominent single gene mutations associated with autism and the animal model research around these genes[2,3]. This leaves a significant gap in evaluating the E:I theory on a larger majority of the autistic population.

While no one theory can fully explain all individuals with an autism diagnosis[4,5], the E:I imbalance theory may have utility for understanding subtypes of autistic individuals[6–8]. Sex/gender may be an important stratifier of relevance for highlighting E:I imbalance subtypes[9,10]. Many highly penetrant autism-associated genes are located on the sex chromosomes (e.g., *FMR1*, *MECP2*, *NLGN3*, *GABRA3*, *ARX*, *SYN1*) and are known to lead to pathophysiology implicating E:I dysregulation[1,11,12]. Other genes playing important roles in the balance between excitation and inhibition in the brain (e.g., *MEF2C, GRIK2, GRIA1, SCN3A, SCN9A, NPTX2*) are highly sensitive to androgens in human neuronal stem cells and are highly expressed in 'social brain' circuitry such as the default mode network, and in particular, the medial prefrontal cortex (MPFC)[13]. Optogenetic stimulation to enhance excitation in mouse MPFC also results in changes in social behavior[14,15]. These results hint that sex-relevant biological mechanisms affect E:I balance and that key social brain regions such as MPFC may be of particular importance for explaining how E:I imbalance affects social behavior.

Sex/gender heterogeneity may also lead to differing clinical presentations and compensatory mechanisms in autism. It is known that many cognitively able adult autistic women engage in camouflaging behaviors that tend to compensate or mask their social-communicative difficulties more so than males[16–18]. Prior work has shown that whereas autistic males show reduced ventral MPFC (vMPFC) self-representation neural response, autistic females show intact vMPFC function. Furthermore, the degree of intact vMPFC self-representation neural activation response in autistic females is associated with enhanced ability to camouflage[19]. If E:I imbalance asymmetrically affects vMPFC function in males versus females, intrinsic E:I neurophysiology could help explain differential camouflaging in adult autistic females.

One possible way to make progress in inferring E:I balance in humans is to develop biomarkers of E:I balance from non-invasive measures of brain activity. Here we pursue the idea that spectral properties of neural time-series data (e.g., local field potentials (LFP) or blood oxygenated level dependent (BOLD) signal) could be used to develop such biomarkers. It has been long known that LFP and rsfMRI data exhibits rich spectral properties, with power decreasing as function of frequency[20–24]. Models of neural networks have reported that the E:I ratio has profound effects on the spectral shape of electrophysiological activity. Recent work with simplified models has proposed that the exponent of the 1/f spectral power law, an index closely related to H, reflects the extent of E:I imbalance[25]. This suggest that neurophysiologically heightened E:I ratio generates flatter 1/f slope in LFP data and this could drive H (as measured in BOLD) to be decreased. In past

work we have shown that a metric related to 1/f slope, the Hurst exponent (H), is atypically decreased in rsfMRI data of adult autistic males, particularly for social brain areas like MPFC[26]. H is statistically relevant to neural noise since lower levels of H can be interpreted as closer to what would be expected of a completely noisy random signal (e.g., white noise produces an H = 0.5). Related to H and long-memory characteristics of the rsfMRI time-series, prior work has also shown case-control differences in the intrinsic neural timescale in autism (e.g., magnitude of temporal autocorrelation)[27]. However, these prior studies examine primarily male-dominated samples and thus cannot shed insight into sex-related heterogeneity in autism.

This work aims to better understand how E:I imbalance may differentially affect autistic males and females. To achieve this aim, we first took a bottom-up approach by using computational models of local neuronal microcircuitry to make *in-silico* predictions about how H and 1/f slope in local field potentials (LFP) and rsfMRI data may behave when there are underlying changes in E:I balance. Importantly, our computational approach takes a major step forward from prior work[25] by utilizing a model that includes interactions within and between excitatory and inhibitory neuronal populations. Our *in-silico* predictions are then tested *in-vivo* with a combination of rsfMRI and chemogenetic manipulations in mice that either increase neurophysiological excitation or that silence the local activity in the network. At the genomic level we then examine what cell types could possibly underlie sex-related heterogeneity in E:I imbalance. Finally, we then turn to the human rsfMRI data to show how E:I imbalance may differ amongst autistic males and females and how such mechanisms may explain individual differences in camouflaging behavior.

Results

Analysis of E:I balance in simulated LFPs from a recurrent network model

In a bottom-up fashion, we first worked to identify potential biomarkers of E:I imbalance from neural time-series data such as local field potentials (LFPs). Motivating our *in-silico* modeling of E:I effects on LFP and BOLD data, we note prior work by Gao and colleagues[25]. This prior work simulated LFP time-series from non-interacting excitatory and inhibitory neuronal populations (Supplementary Figure 1A) and showed that spectral properties such as the 1/f slope flatten with increasing E:I ratio (Supplementary Figure 1B). Given the relationship between 1/f slope and H[28], we show within this modeling approach that as E:I ratio increases, H decreases (Supplementary Figure 1C). However, a limitation of this prior work is it does not include interactions between excitatory and inhibitory populations nor does it allow for recurrent connections within such populations.

With a more biologically plausible recurrent network model of interacting excitatory and inhibitory neuronal populations that receive external inputs (both a sensory driven thalamic input and a sensory unrelated intracortical input) (Figure 1A; see Methods for more details), we manipulated the E:I ratio by independently varying the strengths of the inhibitory (g_I) and excitatory (g_E) synaptic conductances. We called g the relative ratio between inhibitory and excitatory conductances ($g = g_I/g_E$). We report simulation results for two levels of strength of thalamic input (υ_0 = 1.5 spikes/second and υ_0 = 2 spikes/second), and we verified that our results hold qualitatively for a wider range of input levels (1.5 to 4 spikes/second). Figures 1B-C show

examples of LFP time-series and power spectral densities (PSDs) for two values of g, one within an excitation-dominated regime (g = 5.6) and the other within an inhibition-dominated regime (g = 14.8). The spectral profiles (Figure 1C) display two different regions of frequencies with different spectral properties: a region of steeper negative 1/f slopes at higher frequencies (> 30 Hz) and a region of shallower (small negative and sometimes positive) 1/f slopes at low frequencies (< 30 Hz). Thus, we calculated 1/f slopes for the low- and high-frequency regions with piecewise regressions of log power predicted by log frequency. 1/f slopes from the low-frequency (Figure 1D) and high-frequency region (Figure 1E) increase when g is reduced (i.e. E:I ratio augmented). This means that lower values of g correspond to faster spectra with relatively more power at higher frequencies. Changes in 1/f slopes are more prominent in the excitation-dominated region where g is smaller (that is, E:I ratio is shifted in favor of E) than the reference value (g = 11.3), which has been shown to be a plausible reference value that reproduces cortical power spectra well[29–34]. An increase in g beyond this reference value (shifting the E:I balance towards stronger inhibition) had a weaker effect on 1/f slopes. Similar results were obtained when quantifying 1/f slope using the FOOOF algorithm[35] (Supplementary Figure 2), indicating that slopes are not biased by the particular piecewise linear fit procedure. Next, we computed H from the same simulated LFPs. As expected, H decreases with decreasing g (i.e. increasing E:I ratio), but only when g is below the baseline reference value (Figure 1F). These results clearly indicate that, in a biologically plausible computational model of local cortical microcircuitry including recurrent connections between excitatory and inhibitory neuronal populations, changes in synaptic E:I ratio are reflected by and thus could be inferred from the overall LFP readout of 1/f slope or H.

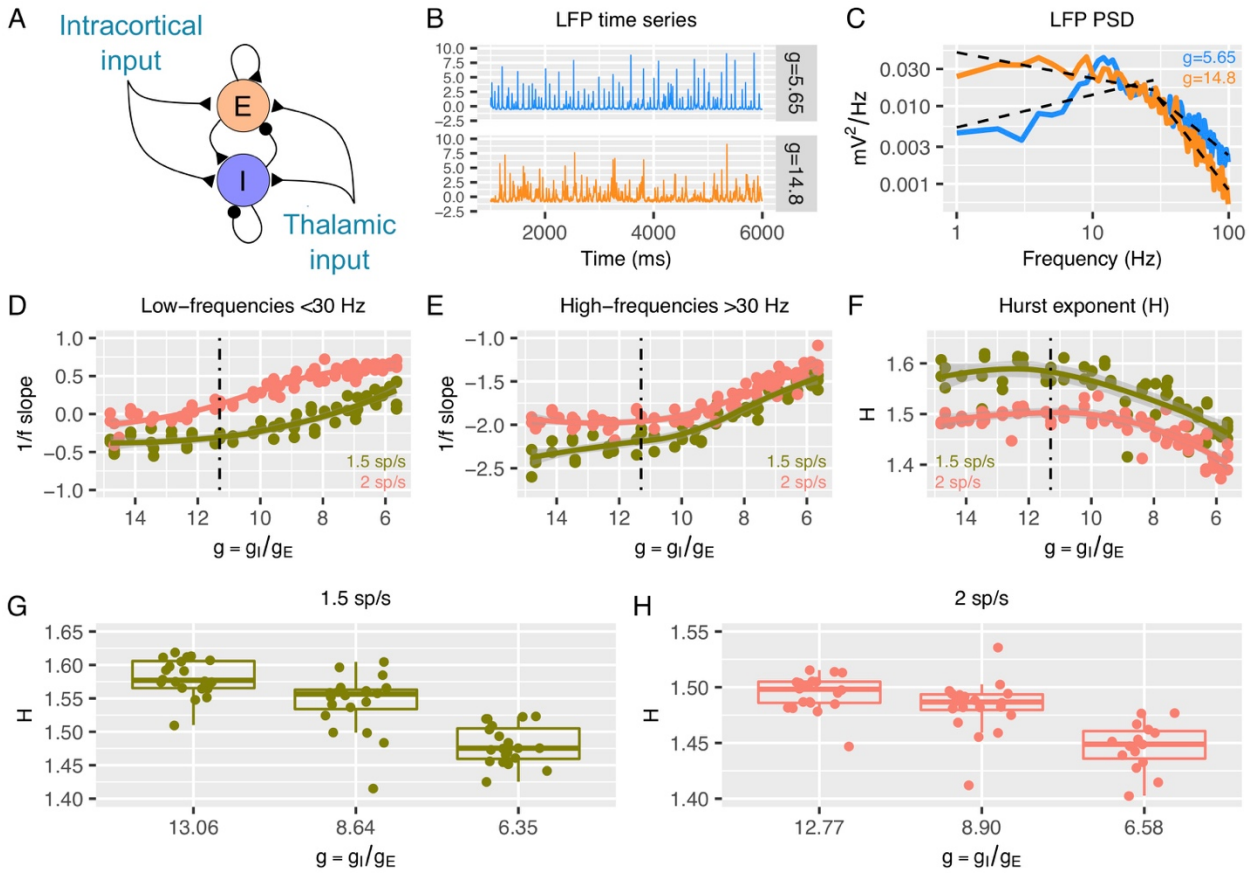


Figure 1: Predictions from a recurrent network model of how the low- and high-frequency slopes of the LFP power spectrum and H change with the variation in relative ratio of inhibitory and excitatory synaptic conductances. *Panel A shows a sketch of the point-neuron network that includes recurrent connections between two types of populations: excitatory cells (E) and inhibitory cells (I). Each population receives two types of external inputs: intracortical activity and thalamic stimulation. Panels B and C show examples of normalized LFP times series and their corresponding PSDs generated for two different ratios between inhibitory and excitatory conductances ($g = g_I/g_E$). The low- and high-frequency slopes of the piecewise regression lines that fit the log-log plot of the LFP PSDs are computed over two different frequency ranges (1-30 Hz for the low-frequency slope and 30 - 100 Hz for the high-frequency slope). The relationship between low-frequency slopes (panel D), high-frequency slopes (panel E) and H values (panel F) are plotted as a function of g for two different firing rates of thalamic input (1.5 and 2 spikes/second). The reference value of g (which has shown in previous studies to reproduce cortical data well) is represented by a dashed black line. In panel G and H, we show H values in 3 different groups of g (high, medium and low g), with the same number of samples in each group.*

Simulated BOLD signal tracks with changes in E:I ratio and correlates selectively with LFP power bands

Given that E:I ratio in LFP data is related to 1/f slope and H, we next asked whether simulated fMRI BOLD signal from the recurrent model would also show similar relationships. To answer this question we first had to simulate BOLD data from the LFP data generated from the recurrent model. Our approach to simulating BOLD (see Methods and Supplementary Figure 3 for how BOLD was simulated from LFP), captures several key characteristics about the empirical relationship between LFP and BOLD. Studies with simultaneous LFP and BOLD measured in non-human primates[36–38] have consistently shown that LFP power show time-lagged correlations with the time course of BOLD signal, and that different frequencies bands vary in how they correlate with measured BOLD signal. In particular, gamma band frequencies tend to show the strongest correlation between LFP power and BOLD signal. Frequency dependency of the EEG-BOLD relationship, with prominent predictive power of the gamma band, is also reported in humans[39]. Remarkably, these empirical observations are recapitulated with simulated LFP and BOLD data from the recurrent model. Figure 2A shows time-lagged correlations between time-dependent LFP power and BOLD. Figure 2B shows that all considered LFP frequency bands (e.g., alpha, beta, gamma) correlate with BOLD, but with gamma band showing the strongest correlations. Thus, our method for simulating BOLD from recurrent model LFP data retains key empirical relationships observed between real LFP power and BOLD. Simulating BOLD with a simple hemodynamic response function (HRF) convolution of the LFP would have not respected the patterns of correlations between LFP power and BOLD observed in empirical data (i.e. the relative increase in correlation between the gamma band and BOLD with respect to other bands; Supplementary Figure 4).

With simulated BOLD from the recurrent model, we next computed H on these data to understand if E:I ratio in the recurrent model is associated with changes in H in BOLD. Strikingly, H in BOLD shows the same dependency on g as observed in LFP data (Figure 2C-D) - H in BOLD decreases as E:I ratio is shifted toward higher excitation by lowering the value of g with respect to the reference value. Although H in LFP and BOLD showed similar associations with respect to

changes in g, it is notable that the range of H in BOLD is shifted towards smaller values (Figure 2C-D) than H in LFP (Figure 1G-H). We also verified that the dependency of H in BOLD on g was largely independent of the details of how BOLD is simulated from LFP. While the results shown in Figure 2 are computed with an HRF that reproduces the correlation function measured between the BOLD signal and the gamma band of LFP[36], it is notable that these results remained similar when using the canonical HRF instead (Supplementary Figure 4). Removing the high pass filter from simulation of BOLD response did alter the relative values of correlation between LFP power and BOLD across frequency bands, making the BOLD response more in disagreement with experimental data, but did not change the relationship of decreasing H with decreasing g (Supplementary Figure 4). In sum, the inferences from the recurrent network model suggest that H in LFP and BOLD data can be utilized as a marker to track changes in underlying synaptic E:I mechanisms.

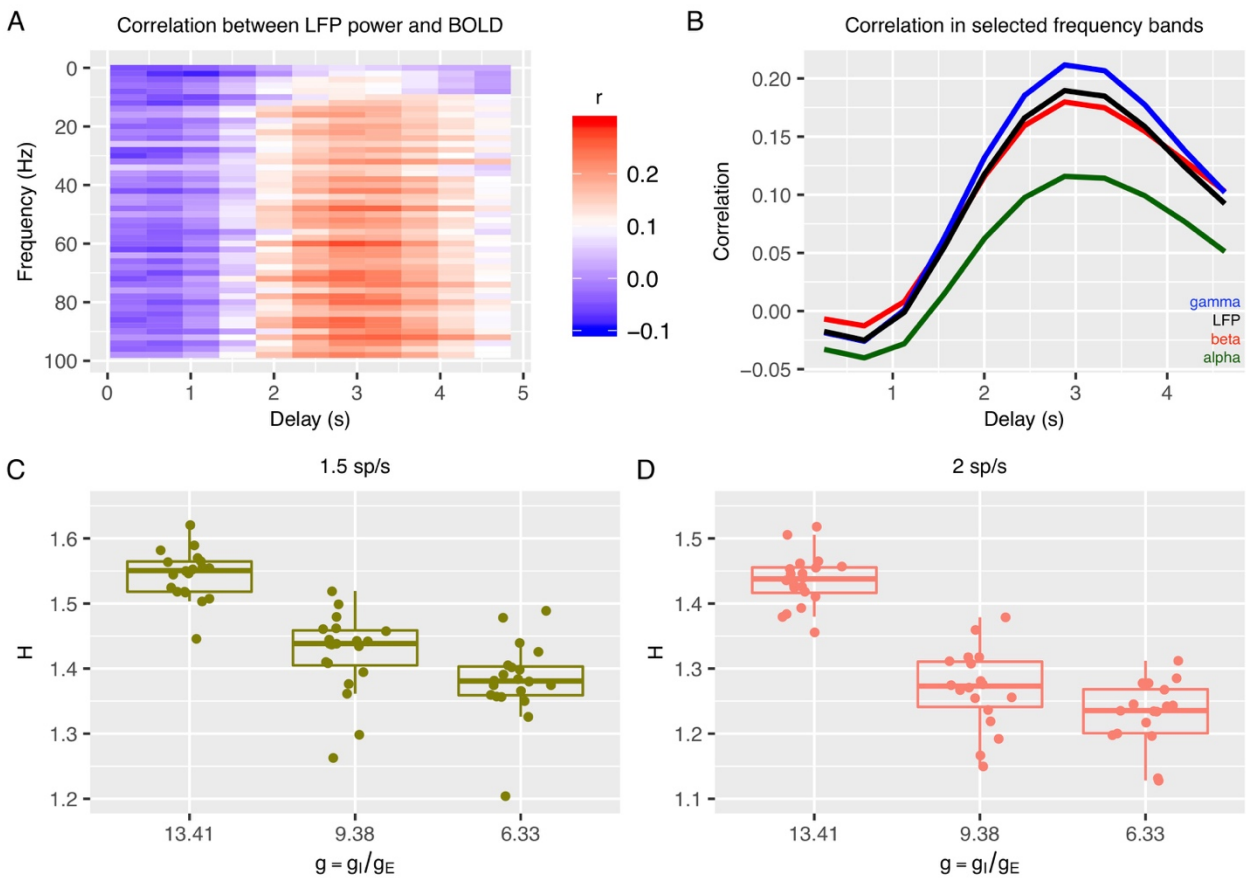


Figure 2: Relationship between E:I ratio from the recurrent model and H measured in simulated BOLD response. *Panel A shows time-lagged Pearson correlations between LFP power across a range of different frequencies and the BOLD signal. Panel B shows the correlation between LFP power and the BOLD signal in selected frequency bands. The following four LFP bands are considered: alpha (8 - 12 Hz), beta (15 - 30 Hz), gamma (40 - 100 Hz) and the total LFP power (0 - 100 Hz). The relationship between E:I ratio (g) and H in simulated BOLD is shown in panels C and D two different firing rates of thalamic input (1.5 and 2 spikes/second).*

Modeling the effects of chemogenetic manipulations within the recurrent network model

We next investigated manipulations of parameters within the recurrent model that approximate the effects of empirical chemogenetic DREADD manipulations in neurons. These simulations are useful to both gain a better understanding of the empirical BOLD measures under DREADD manipulations presented in the next section, and to better characterize the specificity of the origin of changes in 1/f slopes and H with the E:I ratio. Given that, as shown above, in our models changes of H in BOLD mirror those in LFPs, here we present changes in model LFP spectra when simulating these DREADD manipulations.

We first studied the specific effect of solely increasing excitation within the recurrent network. This can be achieved experimentally by using the drug clozapine-N-oxide (CNO) on the DREADD receptor hM3Dq to increase the excitability of excitatory cells only[40]. We simulated this kind of increase of excitability of pyramidal cells in the recurrent network model by lowering their voltage threshold (V_{th}) for spike initiation. Progressively lowering V_{th} from -52 to -53 mV resulted in more positive low-frequency and flatter high-frequency 1/f slopes (Supplementary Figure 5A) and also caused decreases in H (Figure 3A). For H, increasing V_{th} (i.e. decreasing excitability) from -52 to -51 mV resulted in little change in H. These results predict that specific increases of excitation, as in the application of the hM3Dq DREADD to enhance excitability of pyramidal neurons, should reduce steepness of the high-frequency slopes and lead to a decrease in H. These results also confirm our above findings that in recurrent networks in which excitatory and inhibitory neurons interact, increases in excitability are easier to detect from changes in 1/f slope or H than decreases in excitation.

To study whether the changes in 1/f slopes and H are specific to modulations in excitability of only excitatory neurons, we modeled the combined effect of silencing both excitatory and inhibitory neuronal populations. This silencing of both excitatory and inhibitory neurons can be obtained experimentally by application of the hM4Di DREADD (see next Section). In the recurrent network model, we simulated this silencing of both excitatory and inhibitory cells by decreasing the resting potential, E_L, in both excitatory and inhibitory neurons. Decreasing E_L from the baseline value of -70 to -75 mV produced varied effects in 1/f slopes (Supplementary Figure 5B) and resulted in a slight increase of H (Figure 3B). However, these effects were far less prominent than those observed when enhancing excitation specifically (Figure 3A and Supplementary Figure 5A). These results predict a very small, or null, effect of the hM4Di DREADD on H and 1/f slopes. These results also imply that decreases in H are more likely to result from specific increases in excitation rather than from non-specific decreases of excitability across both excitatory and inhibitory neuronal populations.

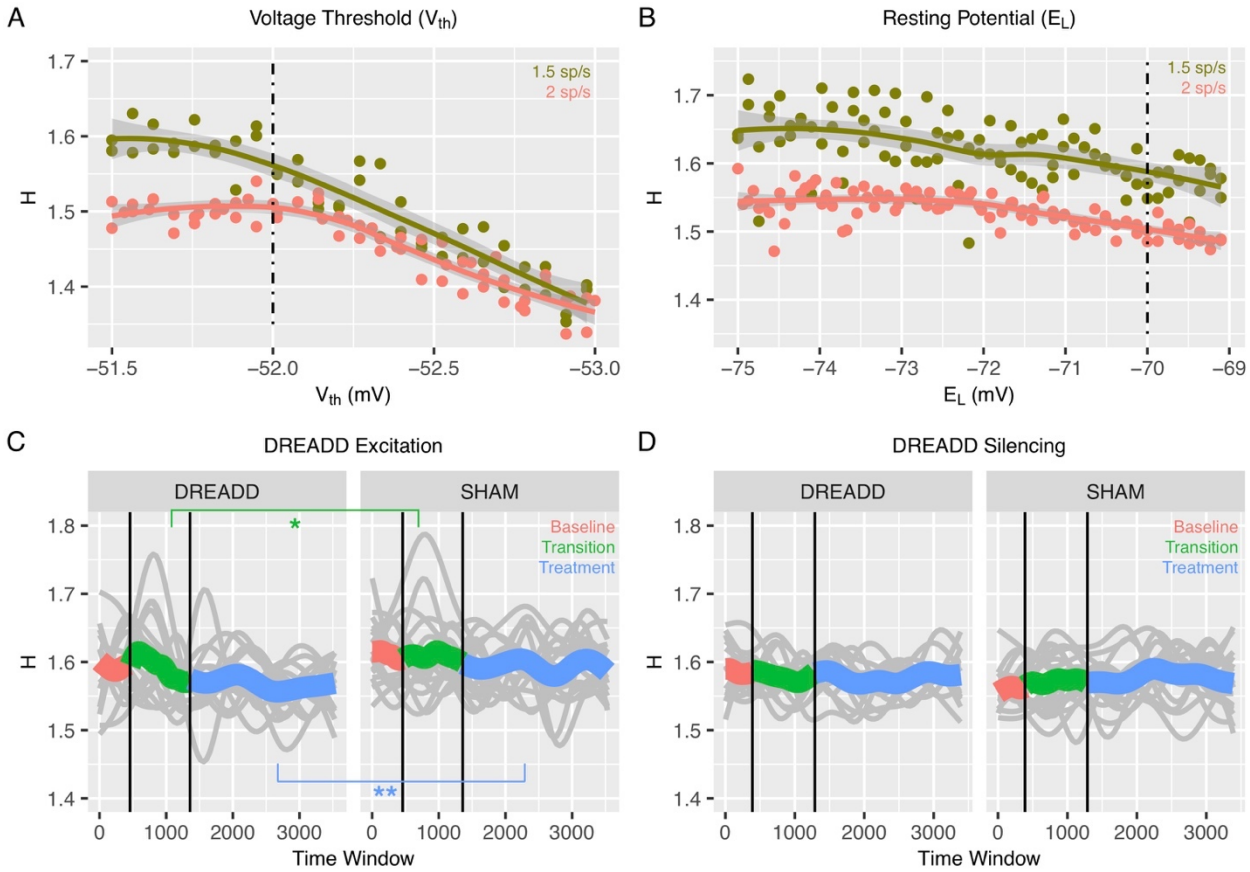


Figure 3: Changes to H after chemogenetic DREADD manipulations to enhance excitability or silence excitatory and inhibitory neurons. *Panel A shows how H changes after the voltage threshold for spike initiation in excitatory neurons (V_{th}) is reduced in the recurrent model, thereby enhancing excitability as would be achieved with hM3Dq DREADD manipulation. Panel B shows how H changes after decreasing the resting potential of excitatory and inhibitory neurons (E_L), as would be achieved with hM4Di DREADD manipulation. the voltage threshold for spike initiation in excitatory neurons (V_{th}) is reduced in the recurrent model, thereby silencing both excitatory and inhibitory neurons as would be achieved with hM3Dq DREADD manipulation. Panel C shows changes in H in BOLD from prefrontal cortex (PFC) after real hM3Dq DREADD manipulation in mice, while panel D shows changes in PFC BOLD H after hM4Di DREADD manipulation in mice. In panels C and D, individual gray lines indicate H for individual mice, while the colored lines indicate Baseline (pink), Transition (green), and Treatment (blue) periods of the experiment. During the Baseline period H is measured before the drug or SHAM injection is implemented. The Transition phase is the period after injection but before the period where the drug has maximal effect. The Treatment phase occurs when the drug begins to exert its maximal effect. The green star indicates a Condition*Time interaction ($p<0.05$) in the Transition phase, whereas the blue stars indicate a main effect of Condition within the Treatment phase ($p < 0.005$).*

Changes in H in BOLD after chemogenetic manipulation to enhance excitability of excitatory neurons in mice

The *in-silico* modeling presented thus far makes the prediction that in real BOLD data, if E:I ratio is increased via enhanced excitability of excitatory neurons, then H should decrease. To

empirically test this prediction *in-vivo*, we measured rsfMRI BOLD signal in prefrontal cortex (PFC) of mice under conditions where a chemogenetic manipulation (hM3Dq DREADD)[40] is used to enhance excitability of pyramidal neurons. Here we used a sliding window analysis to assess dynamic changes in H over the course of 3 different phases of the experiment – 1) a 'Baseline' phase where the CNO drug or a SHAM injection had not yet occurred, 2) a 'Transition' phase directly following CNO or SHAM injection, and 3) a 'Treatment' phase, whereby CNO has its maximal effect. We find that H is modulated over time by the DREADD manipulation (condition*time*treatment phase interaction $F = 349.03$, $p < 0.0001$). During the Baseline phase of rsfMRI scanning before the DREADD-actuator clozapine N-oxide (CNO) was injected, H under DREADD or a SHAM control conditions are not affected (condition main effect $F = 0.82$, $p = 0.37$; condition*time interaction $F = 0.36$, $p = 0.54$). However, during the Transition phase of the experiment where the CNO begins to have its effects, we find a condition*time interaction ($F = 4.94$, $p = 0.0262$), whereby H drops over time at a steeper rate during the DREADD condition compared to the SHAM condition (green line in Figure 3C). Finally, during the Treatment phase of the experiment, where the drug exerts its maximal effect, there is a significant main effect of condition ($F = 12.92$, $p = 0.0011$) and no condition*time interaction ($F = 0.66$, $p = 0.4182$) (blue line in Figure 3C) (Table 1). This effect is explained by H being reduced in the DREADD vs SHAM condition. These *in-vivo* results are directly in line with the *in-silico* prediction that enhancing E:I ratio via enhancing the excitability of excitatory neurons results in a decrease in H (i.e. Figure 1F-H, Figure 2C-D, and Figure 3A).

Chemogenetically silencing both excitatory and inhibitory neurons has no effect on H in BOLD

While the above results show that specific enhancement of excitability in excitatory neurons results in a decrease in BOLD H, it is an important negative control contrast to investigate whether non-specifically reducing the excitability of both excitatory and inhibitory neuronal populations might also affect H. This is an important negative control since if H were to change in a similar direction after this manipulation, it would make interpretations about decrease in H being due to increased E:I ratio via excitation problematic. The *in-silico* simulation of this manipulation (Figure 3B) would predict that H would not be changed much, and that if a change in H were to occur, it would be a slight increase rather than a decrease in H. By expressing the inhibitory hM4Di DREADD[41] under the control of a pan-neuronal promoter, we chemogenetically silenced both excitatory and inhibitory neurons in PFC of mice and re-ran the same rsfMRI neuroimaging protocol as before. While a significant 3-way interaction between condition, time, and treatment phase was present ($F = 85.8$, $p < 0.0001$), there were no strong main effects of condition or condition*time interactions in any of the baseline, transition, or treatment phases of the experiment (see Table 2 and Figure 3D). Overall, these results along with the recurrent model simulation of h4MDi DREADD (Figure 3B) bolster strength of the interpretation that enhanced excitation drives decreasing H in BOLD and that H in BOLD would not change appreciably in a situation such as pan-neuronal silencing of both excitatory and inhibitory neurons.

Autism-associated genes in excitatory neuronal cell types in the human brain are enriched for genes that are differentially expressed by androgen hormones

The findings thus far suggest that excitation affects metrics computed on neural time-series data such as 1/f slope and H. Applied to the idea of sex-related heterogeneity in E:I imbalance in autism, these results make the prediction that excitatory neuronal cell types would be the central cell type affecting such neuroimaging phenotypes in a sex-specific manner. To test this hypothesis about sex-specific effects on excitatory neuronal cell types, we examined whether known autism-associated genes that affect excitatory neuronal cell types[42,43] are highly overlapping with differentially expressed genes in human neuronal stem cells when treated with a potent androgen hormone, dihydrotestosterone (DHT)[13,44]. Genes differentially expressed by DHT are highly prominent within the gene set of autism-associated genes that affect excitatory neurons (OR = 1.67, p = 0.03), with most of the overlapping genes being those whereby DHT upregulates expression (Figure 4A). By contrast, genes associated with autism that affect inhibitory neuronal cell types or other non-neuronal cells (e.g., microglia, astrocytes, oligdendrocytes) are not enriched for DHT differentially expressed genes (inhibitory neurons: OR = 1.51, p = 0.12; microglia: OR = 0.78, p = 0.78; astrocytes or oligodendrocytes: OR = 1.11, p = 0.49). This result suggests that autism-associated genes specifically affecting excitatory neuronal cell types are also susceptible to the male-specific influence of androgen hormones in human neuronal stem cells.

We next additionally examined how such DHT-sensitive and autism-associated excitatory neuron genes spatially express in the human adult brain. This analysis would help shed insight on which brain areas might be more affected by such sex-specific effects in autism. A one-sample t-test of gene maps from the Allen Institute Human Brain Atlas[45] shows that this subset of DHT-sensitive and autism-associated excitatory neuron genes are highly expressed in MPFC, PCC, and anterior insula, amongst many other areas (Figure 4B-C).

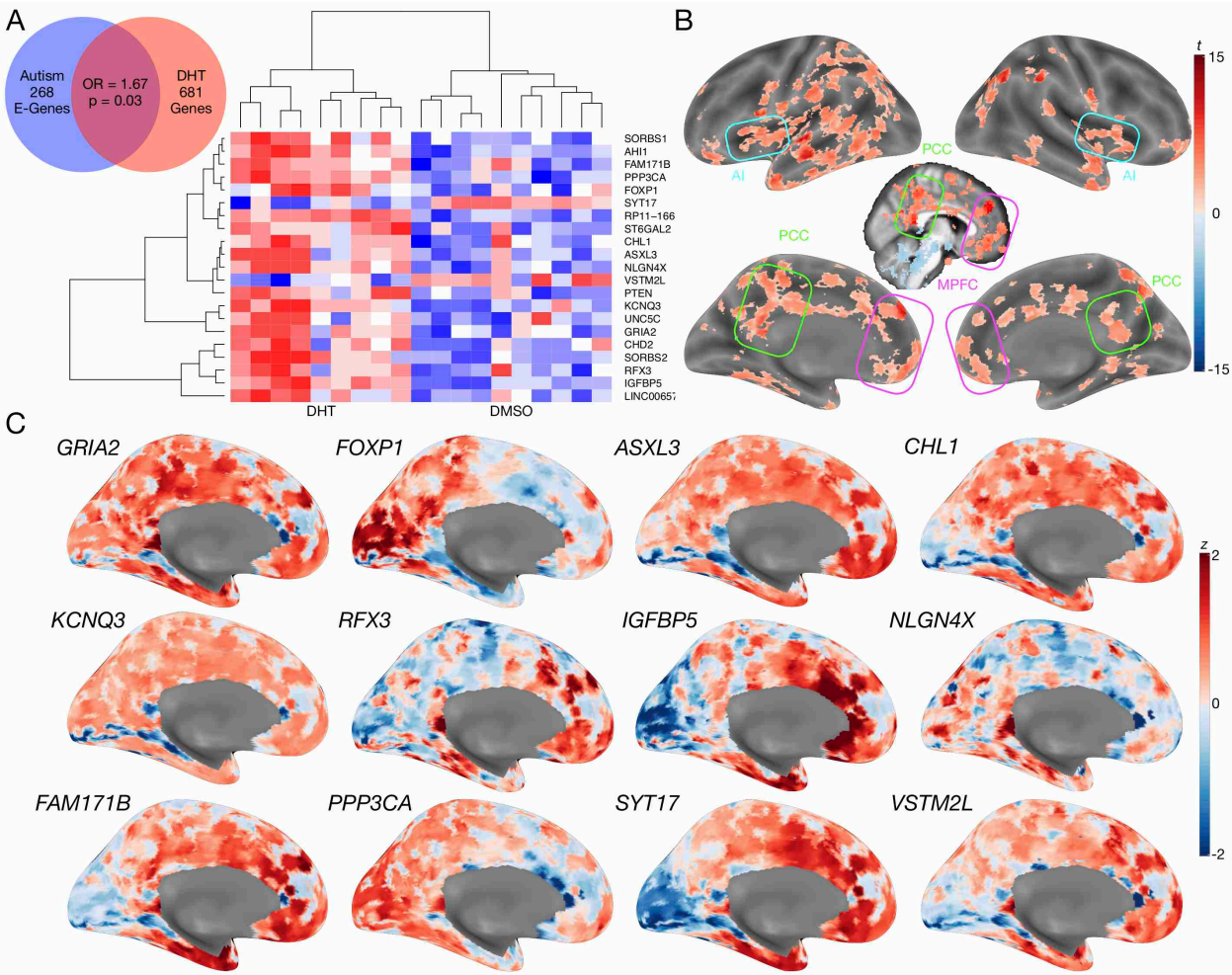


Figure 4: Autism-associated genes within excitatory neuronal cell types are enriched for genes differentially expressed by androgen hormones. *Panel A shows a Venn diagram depicting the enrichment between autism-associated genes affected excitatory neurons and DHT-sensitive genes. Panel A also includes a heatmap of these genes whereby the color indicates z-normalized expression values. The column dendrogram clearly shows that all samples with DHT treatment are clustered separately from the control DMSO samples. Each row depicts the expression of a different gene. Panel B shows a t-statistic map from a whole-brain one-sample t-test on these DHT-sensitive and autism-associated genes in excitatory neurons. Results are thresholded at FDR q<0.01. Panel C shows spatial gene expression profiles on a representative surface rendering of the medial wall of the cortex for specific genes shown in panel B. Each map shows expression as z-scores with the color scaling set to a range of -2<z<2.*

H is on-average reduced in adult autistic men but not women

We next move to application of this work to human rsfMRI data in adult autistic men and women. If E:I ratio is affected by sex-related mechanisms[13], we predict that H would be differentially affected in autistic males versus females and manifest as sex-by-diagnosis interactions in a 2x2 factorial design (Sex: Male vs Female; Diagnosis: Autism vs Typically-Developing (TD)). More specifically, the directionality of our predictions from the *in-silico* and *in-vivo* results in Figures 1-3 are that if H reflects E:I ratio, there should be decreased H (due to

enhanced E) specifically in autistic males but not autistic females. Mass-univariate analysis uncovered one region in ventromedial prefrontal cortex (vMPFC) with a sex-by-diagnosis interaction passing FDR q<0.05 ($F(5,104)$ = 15.13, p = 0.0001, *partial* η^2 = 0.12) (Figure 5A). In line with directionality of our predictions, this interaction effect is driven by a large TD>Autism effect in males (*Cohen's d* = 1.30) and a small Autism>TD effect in females (*Cohen's d* = -0.27) (Figure 5B). A similar sex*diagnosis interaction appears when using another metric such as the intrinsic neural timescale[27] (Supplementary Figure 6). We also used a multivariate partial least squares (PLS) analysis to uncover that vMPFC is part of a more distributed neural system expressing the same sex-by-diagnosis interaction (d = 2.04, p = 0.036) as other areas that include relevant default mode network (DMN) areas in posterior cingulate cortex/precuneus (PCC) (Supplementary Figure 6), and other non-DMN areas such as anterior insula, lateral prefrontal cortex, somatosensory and motor cortices, amongst others (Figure 5C). The PLS result allows for detection of other regions that were subthreshold in the mass-univariate analysis, but which showed heightened effect sizes (e.g., white and light blue areas in the unthresholded map shown in Figure 5A). Detection of these regions in a mass-univariate analysis may require a larger sample size to enhance statistical power. While the main effects of diagnosis and sex are not the primary contrast for this study, we report that no significant regions survived FDR q<0.05 for the main effects of diagnosis. However, 61% of brain regions showed an on-average male>female sex difference (Supplementary Figure 6), which is within keeping of results from other work on sex differences in H[46].

Correlation between vMPFC H and camouflaging in autistic women but not men

In prior task-fMRI work we found a similar sex-by-diagnosis interaction in vMPFC self-representation response and a female-specific brain-behavioral correlation with camouflaging ability[19]. Given that adult autistic females engage more in camouflaging on-average[16–18], we next asked whether vMPFC H would be related to camouflaging in a sex-specific manner. In autistic females, increased camouflaging was strongly associated with increased H in vMPFC (r = 0.60, p = 0.001). However, no significant association was apparent in autistic males (r = -0.10, p = 0.63). The strength of this brain-behavioral correlation significantly differed between autistic males and females (z = 2.58, p = 0.009) (Figure 5D). This result suggests that the more intact H values in vMPFC of autistic females, which are likely reflective of more intact E:I balance, are linked to better ability to camouflage social-communicative difficulties typically seen in autistic females.

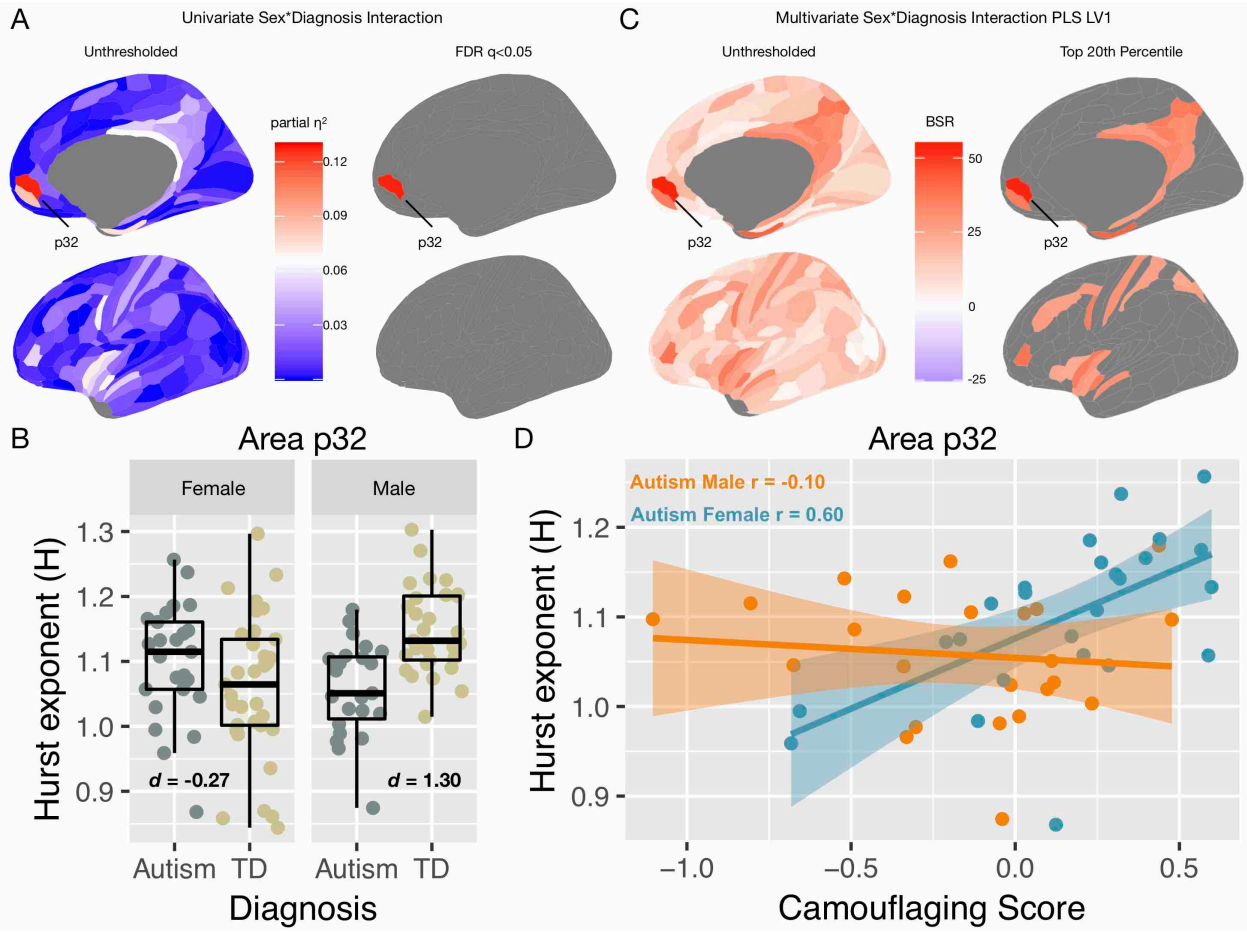


Figure 5: Autism rsfMRI sex*diagnosis interaction results. *Panel A shows unthresholded and thresholded with FDR q<0.05 mass-univariate results for the sex*diagnosis interaction contrast. Panel B shows H estimates from vMPFC (area p32) across males and females with and without autism. Panel C shows partial least squares (PLS) results unthresholded and thresholded to show the top 20% of brain regions ranked by bootstrap ratio (BSR). Panel D shows correlation between vMPFC H and behavioral camouflaging score in autistic males (orange) and females (blue).*

Discussion

In this work we set out to better understand how intrinsic E:I imbalance affects the autistic brain in a sex-specific manner. Evidence from animal models of rare genetic variants associated with autism have typically been used as the primary evidence for the E:I imbalance theory[1,2]. However, these variants affect only a small percentage of the autism population. Thus, it is unclear how E:I imbalance might affect the majority of heterogeneous individuals within the total autism population. To bridge this gap we need multi-level methods that can be applied to understand the 'living biology' behind actual human individuals[47], such as in-vivo neuroimaging data and metrics applied to such time-series data that are linked to actual underlying neural E:I mechanisms. Bridging this gap will help us identify mechanistic targets that explain neural and behavioral variability across a much larger portion of individuals in the autism population.

Based on earlier work[25], we reasoned that metrics such as 1/f slope and H in neural time-series data would be relevant as an *in-vivo* neuroimaging marker of E:I mechanisms. Prior work suggested this relationship via a model that considers inhibition and excitation as separate entities[25]. However, excitation and inhibition in the brain are inseparably linked. Furthermore, prior work[25] considered only 1/f slopes in simulated LFP data and did not explore the effect of the transformation between LFP neural activity to BOLD. Our simulations address these problems and significantly extend prior work[25] on the relationship between E:I imbalance and changes in spectral properties of neural signals. We showed that when excitation and inhibition interact in a recurrent network model, flatter 1/f slopes and decreases in H are specific markers of increases in E:I ratio. We also showed that in simulated BOLD signal, H and E:I ratio are associated in a manner similar to the relationships observed with LFP data. Taken together, these results predict that changes in H in neural time-series data can be interpreted as a shift in synaptic E:I ratio that permeates through in LFP or BOLD readouts.

The power of our *in-silico* modeling approach is that it provides explicit predictions of what to expect in real BOLD data when synaptic E:I imbalance occurs. Remarkably, these *in-silico* predictions are confirmed *in-vivo* with rsfMRI BOLD data in mice after experimental chemogenetic manipulations that specifically enhance neural excitation. Intriguingly, and consistent with *in-silico* predictions, manipulations that silence both excitatory and inhibitory neuronal populations do not have a strong effect on H in BOLD. These results are in line with optogenetic studies showing that specifically enhancing excitation in MPFC seems to have the biggest effects on social behavior in mice[14]. The present work clearly shows that enhancement of excitation results in measurable changes in BOLD readouts as decreases in H. This insight allows us to leverage H as an *in-vivo* rsfMRI biomarker that has strong relevance back to synaptic E:I imbalance.

With regards to how sex-related heterogeneity in E:I imbalance might manifest in autism, we first utilized genomics data and found that autism-associated genes that affect excitatory neuronal cell types are enriched for genes that are differentially expressed by DHT. This inference extends prior work implicating excitatory neuron cell types in autism-relevant biology[42,43,48,49] by linking genomic mechanisms in these cell types to male-specific influence of androgen hormones. Importantly, other cell types such as inhibitory neurons do not express autism-associated genes that are also influenced by DHT. Additionally, the DHT-sensitive and autism-associated excitatory genes tend to spatially express in the human adult brain in regions such as MPFC, PCC and anterior insula, which have been shown to be affected across a range of task-related and rsfMRI studies[19,50–54].

Moving to human rsfMRI data on adult patients with autism, we utilized H as a neuroimaging biomarker of E:I imbalance. Specifically, we examined whether H differs between adult males and females with and without autism. Mass-univariate analysis highlighted one region, vMPFC, which showed a sex-by-diagnosis interaction - that is, H was specifically reduced in adult autistic males, but not in autistic females. Reduced H in autistic males is compatible with the inference of elevated E:I ratio potentially driven by enhanced excitation. Multivariate PLS analysis extended this finding by showing that a distributed neural system structurally and functionally connected to vMPFC, such as default mode network (DMN) areas like PCC[55,56], and anterior insula[52], also expressed a similar but more subtle sex-by-diagnosis interaction. Interestingly, these

regions highlighted by the PLS analysis are remarkably similar to the map of brain regions where autism-associated excitatory and DHT-sensitive genes highly express (Figure 4B-C). Therefore, important social brain circuitry such as the DMN, and other integrative hubs of the salience network (e.g., anterior insula) that connect DMN to other important large-scale networks[52] may be asymmetrically affected by heightened E:I ratio in autistic males more than autistic females.

These human rsfMRI results are not only compatible with the *in-silico* predictions and the *in-vivo* mouse rsfMRI data presented here, but are also compatible with several prior lines of work. Our prior work highlighted that DMN functional connectivity in adolescent males, but not females, is affected by heightened levels of fetal testosterone and this network was heavily comprised of MPFC and PCC[13]. In the same work, we showed that a cortical midline DMN subsystem comprising MPFC and PCC highly expresses several genes relevant for excitatory postsynaptic potentials (e.g., *MEF2C, GRIK2, GRIA1, SCN3A, SCN9A, NPTX2*). The current findings linking autism-associated genes in excitatory neuron cell types (Figure 4) dovetails with and extends these findings with specific inferences about the importance of excitatory cell types over and above other inhibitory cell types, where evidence about their role in E:I imbalance in autism is more mixed (e.g., [57,58]). Importantly, the expression of these genes in human neuronal stem cells are elevated after exposure to the potent androgen DHT[13]. Thus, one potential explanation for the male-specific reduction of H in vMPFC could have to do with early developmental and androgen-sensitive upregulation of genes that play central roles in excitatory neuron cell types, and thus ultimately affecting downstream E:I imbalance. Such effects may be sex-differential and thus less critical in human females, serving an important basis of sex-differential human brain development[59] and explaining the sex-based heterogeneity and qualitative sex differences of autism neurobiology in human[10,60].

rsfMRI H in autistic adults was also relevant in a sex-specific manner to a clinical behavioral phenomenon known as 'camouflaging'. Camouflaging relates to a set of compensatory or masking strategies/mechanisms that allow individuals to cope with their social-communicative difficulties in everyday social situations[16,17,61]. It is known that cognitively able adult autistic females tend to engage in more camouflaging behavior than males[16–18] and the extent to which individual females engage in camouflaging is linked to vMPFC function[19]. One of the most important known functions of vMPFC has to do with self-representation[54] and simulating others based on information about the self[62]. In prior task-related fMRI work we found a similar sex-by-diagnosis interaction effect whereby males are more impaired in vMPFC self-representation response than their female autistic counterparts. Furthermore, increased magnitude of vMPFC self-representation neural response correlates with increased camouflaging ability, but only in adult autistic females[19]. Strikingly, here we find a similar sex-by-diagnosis interaction effect in vMPFC H as well as a female-specific correlation with camouflaging - as vMPFC H increases, indicative of a more normative or intact level of E:I balance, camouflaging also increases. This converging set of results provides strong evidence that intrinsic mechanisms such as E:I balance may be atypical only in cognitively able autistic males at vMPFC. vMPFC in similarly cognitively able autistic females likely has more intact, or normalized, intrinsic E:I balance and this more intact mechanism potentially allows for intact levels of neural self-representation and enhanced camouflaging ability. Future work changing E:I balance in vMPFC may provide a useful avenue for ameliorating daily life social-communication adaptation and coping difficulties in autistic males and enable them to optimally engage in compensatory processes such as camouflaging to

the similar extent as autistic females. It may also be fruitful to examine how intact E:I balance in vMPFC of females may be an expression of protective factors that are hypothesized to buffer risk for autism in females[63,64].

In conclusion, we show that spectral properties of neural time-series data, such as H and 1/f slope, can be utilized in neuroimaging readouts like LFP and BOLD as a biomarker for underlying E:I-relevant mechanisms. *In-silico* predictions from simulated LFP and BOLD data were confirmed *in-vivo* with rsfMRI BOLD data where excitation was enhanced through chemogenetic manipulation. Finally, in application to humans, we show that H in rsfMRI data is reduced in vMPFC and other DMN areas of adult autistic males, but not females. Reduced H is indicative of enhanced excitation and thus points to sex-specific dysregulation of E:I balance in social brain networks of autistic males. This male-specific dysregulation of E:I balance may be linked to sex-differential early developmental events such as androgen-upregulation of gene expression for genes that play important roles in excitatory postsynaptic potentials[13]. The intact levels of H in females may help facilitate elevated levels of compensation known as camouflaging to cope with daily social-communicative difficulties. This important female-specific brain-behavioral correlation may also be key for future innovation in intervention targeting E:I mechanisms and MPFC-related brain regions that may enable better coping with daily social-communicative adaptation difficulties related to autism. More generally, this work extends the relevance of the E:I imbalance theory of autism beyond evidence from autism-associated rare genetic variants and specify a larger portion of the autism population whereby these E:I mechanisms may be of critical importance.

Methods

Human Participants

All procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All human participants' informed consent was obtained in accord with procedures approved by the Suffolk Local Research Ethics Committee. Adult native English speakers (n=136) with normal/corrected-to-normal vision participated: n=33 typically developing (TD) males, n=34 autistic males, n=34 TD females and n=34 autistic females (Table 3). They all reported cis-gender identity based on a single item inquiring their birth-assigned sex and another on their identified gender. Groups were not statistically different on age or full-scale IQ (FIQ) on the Wechsler Abbreviated Scales of Intelligence (WASI) (Table 3). Exclusion criteria for all participants included a history of or current psychotic disorders, substance-use disorders, severe head injury, genetic disorders associated with autism (e.g. fragile X syndrome and tuberous sclerosis), intellectual disability (i.e. Full-scale IQ (FIQ) < 70), or other medical conditions significantly affecting brain function (e.g. epilepsy).

The inclusion criterion for both male and female autistic participants was a formal clinical diagnosis of International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) childhood autism or Asperger's syndrome, or Diagnostic and Statistical Manual of Mental Disorders (4th ed., text rev.; DSM-IV-TR) autistic disorder or Asperger's disorder, as assessed by a psychiatrist or clinical psychologist in the National Health Service, UK. Since all participants were adults, we further considered available information of developmental history to include only those with clinically evident childhood autistic symptoms, for example, from information collected using the Autism Diagnostic Interview–Revised (ADI-R)[65] where possible, or from the participants' clinical diagnosis letters shared with the research team to determine eligibility. We used this clinically based criterion for inclusion for the purpose of sampling autistic individuals currently diagnosed by specialists in mental health services in the daily practice and to align with best clinical practice as recommended by the UK National Institute for Health and Clinical Excellence (NICE) guideline[66]. For assessing levels of autism characteristics, we administered the Autism Spectrum Quotient (AQ)[67], module 4 of the Autism Diagnostic Observation Schedule (ADOS)[68], and ADI-R[65] where possible, before the fMRI session. Autistic male and female groups were not different on ADI-R Reciprocal-Social-Interaction scores or Reading the Mind in the Eyes Test (RMET)[69] performance (Table 3).

We further used criteria for inclusion based on characteristics about data quality (see next paragraphs for data preprocessing). In particular, we excluded participants where the number of volumes was not acquired due to scanner hardware issues (n=1), the preprocessing pipeline could not adequately preprocess the data (e.g., bad registrations; n=5). Participants were also excluded if their head motion exceed a mean framewise displacement (meanFD) of >0.4mm (n=8). For the remaining subjects we further visually inspected plots of framewise displacement (FD) and DVARS[70] traces to determine whether the wavelet despiking step sufficiently attenuated artefact-related variability that would leave DVARS spikes. Here we made a qualitative and consensus judgement amongst authors (S.T. and M.V.L) to exclude individuals (n=9) whereby there were numerous FD spikes above 0.5mm or numerous DVARS spikes leftover after wavelet despiking

was applied. Other exclusions included any VIQ or PIQ <70 (n=1) and co-morbid agenesis of the corpus callosum (n=1). The final sample sizes included in all further analyses was n=29 TD males, n=23 autistic males, n=33 TD females, and n=25 autistic females. The final groups used in all analyses did not statistically differ on age (Diagnosis main effect: $F(3,106) = 0.03$, $p = 0.85$; Sex main effect: $F(3,106) = 0.14$, $p = 0.70$; Sex*Diagnosis interaction: $F(3,106) = 0.25$, $p = 0.61$) or FIQ (Diagnosis main effect: $F(3,106) = 3.38$, $p = 0.07$; Sex main effect: $F(3,106) = 0.48$, $p = 0.48$; Sex*Diagnosis interaction: $F(3,106) = 2.24$, $p = 0.13$) (see Table 3).

Human fMRI data acquisition

Imaging was performed on a 3T GE Signa Scanner at the Cambridge Magnetic Resonance Imaging and Spectroscopy Unit. Participants were asked to lie quietly in the scanner awake with eyes closed for 13 minutes and 39 seconds during sequential acquisition of 625 whole-brain T2*-weighted echo planar image volumes with the following parameters: relaxation time = 1302 ms; echo time = 30 ms; flip angle = 70°; matrix size = 64 x 64; field of view = 24 cm; 22 anterior commissure-posterior commissure aligned slices per image volume; 4 mm axial slice thickness; 1 mm slice gap. The first five time-points were discarded to allow for T2-stabilization. During analysis of the Hurst exponent (H) for BOLD time-series, due to the discrete wavelet transform using volumes in power of 2, only the first 512 volumes (2^9) were utilized. A high-resolution spoiled gradient anatomical image was acquired for each participant for registration purposes.

Human fMRI data analysis

Preprocessing of the resting state data was split into two components; core preprocessing and denoising. Core preprocessing was implemented with AFNI[71] (http://afni.nimh.nih.gov/) using the tool speedypp.py (http://bit.ly/23u2vZp)[72]. This core preprocessing pipeline included the following steps: (i) slice acquisition correction using heptic (7th order) Lagrange polynomial interpolation; (ii) rigid-body head movement correction to the first frame of data, using quintic (5th order) polynomial interpolation to estimate the realignment parameters (3 displacements and 3 rotations); (iii) obliquity transform to the structural image; (iv) affine co-registration to the skull-stripped structural image using a gray matter mask; (v) nonlinear warping to MNI space (MNI152 template) with AFNI 3dQwarp; (v) spatial smoothing (6 mm FWHM); and (vi) a within-run intensity normalization to a whole-brain median of 1000. Core preprocessing was followed by denoising steps to further remove motion-related and other artifacts. Denoising steps included: (vii) wavelet time series despiking ('wavelet denoising'); (viii) confound signal regression including the 6 motion parameters estimated in (ii), their first order temporal derivatives, and ventricular cerebrospinal fluid (CSF) signal (referred to as 13-parameter regression). The wavelet denoising method has been shown to mitigate substantial spatial and temporal heterogeneity in motion-related artifact that manifests linearly or non-linearly and can do so without the need for data scrubbing[73]. Data scrubbing (i.e. volume censoring) cannot be used in our time-series-based analyses here as such a procedure breaks up the temporal structure of the time-series in such a way that invalidates estimation of the Hurst exponent (H) that examine long-memory characteristics. Wavelet denoising is implemented with the Brain Wavelet toolbox (http://www.brainwavelet.org). The 13-parameter regression of motion and CSF signals was achieved using AFNI 3dBandpass with the –ort argument. To further characterize motion and its impact on the data, we computed FD and DVARS[70]. Between-group comparisons showed that all groups were similar with respect

to head motion as measured by meanFD with no Diagnosis ($F(3,106) = 1.77$, $p = 0.18$) or Sex ($F(3,106) = 0.51$, $p = 0.47$) main effects or Sex*Diagnosis interaction ($F(3,106) = 1.10$, $p = 0.29$). All groups showed average meanFD of less than 0.2 mm (see Table 3).

Mean time-series for each of the 180 parcels within the Human Connectome Project Multimodal Parcellation (HCP-MMP)[74] were extracted from the final preprocessed data, to estimate H. The estimation of H utilizes a discrete wavelet transform and a model of the time-series as fractionally integrated processes (FIP) and is estimated using maximum likelihood estimation. This method utilizing the FIP model for estimating H differs from our prior work[26], which used a model of fractional Gaussian noise (fGn). fGn is one type of process subsumed under the FIP model. However, the fGn model has the limitation of assuming that the BOLD time-series is stationary and also limits the upper bound of H at 1. In practice, we have seen that the upper bound of H=1 from the fGn model results in ceiling effects for many brain regions and subjects. Thus, to remove the assumption of stationarity and upper bound of H=1, the FIP model offers more flexibility and potentially added sensitivity due to better estimation of between-subject variability when estimates are near or exceed H=1. When H>1 the time-series is considered non-stationary and has long memory characteristics (e.g., is fractal). H is computed using the nonfractal MATLAB toolbox written by one of the co-authors (WY) (https://github.com/wonsang/nonfractal). The specific function utilized is bfn_mfin_ml.m function with the 'filter' argument set to 'haar' and the 'ub' and 'lb' arguments set to [1.5,10] and [-0.5,0], respectively.

After H was estimated for each of the 180 HCP-MMP parcels, we used a general linear model to test for Sex*Diagnosis interactions as well as main effects of Sex and Diagnosis in H. These models also incorporated meanFD and FIQ as covariates of no interest. Multiple comparison correction was achieved using an FDR q<0.05 threshold. Visualization of effect sizes for figures was achieved using the ggseg library in R (https://github.com/LCBC-UiO/ggseg).

In addition to mass-univariate analysis, we also utilized multivariate partial least squares (PLS) analysis[75] to highlight distributed neural systems that capture the effect of a sex*diagnosis interaction. This analysis was implemented with code from the plsgui MATLAB toolbox (http://www.rotman-baycrest.on.ca/pls/). A matrix with participants along the rows and all 180 HCP-MMP parcels along with columns was input as the primary neuroimaging matrix for PLS. We also inserted a vector describing the sex*diagnosis contrast as the matrix to relate to the neuroimaging matrix. This vector describing the sex*diagnosis interaction was computed by matrix multiplication of the contrast vector of [1, -1, -1, 1] to a design matrix that was set up with columns defining TD males, autism males, TD females, and autism females, respectively. The PLS analysis was run with 10,000 permutations to compute p-values for each latent-variable (LV) pair and 10,000 bootstrap resamples in order to compute bootstrap ratios (BSR) to identify brain regions of importance for each LV pair. To isolate specific brain regions of importance for a statistically significant LV, we selected the top 20th percentile of brain regions ranked by BSR.

Relationships between H and camouflaging were conducted within autistic males and females separately. Pearson's correlations were used to estimate the strength of the relationship and groups were compared on the strength of the relationship using Fisher's r-to-z transform as implemented with the paired.r function in the *psych* library in R.

Behavioral index of camouflaging

Camouflaging (consciously or unconsciously compensating for and/or masking difficulties in social–interpersonal situations) was operationalized as prior work[16,19]: the discrepancy between extrinsic behavioral presentation in social–interpersonal contexts and the person's intrinsic status. We used both the AQ score and RMET correct score as reflecting intrinsic status (i.e. self-rated dispositional traits and performance-based socio-cognitive/mentalizing capability), and the ADOS Social-Communication total score as reflecting extrinsic behavioral presentation. The three scores were first standardized (S_{ADOS}, S_{AQ} and S_{RMET}) within our sample of autistic men and women by mean-centering (to the whole autism sample in this study) and scaling (i.e. divided by the maximum possible score of each) to generate uniformly scaled measures that can be arithmetically manipulated. The first estimate of camouflaging was quantified as the difference between self-rated autistic traits and extrinsic behaviors ($CF1 = S_{AQ} - S_{ADOS}$), and the second estimate between mentalizing ability and extrinsic behaviors ($CF2 = -S_{RMET} - S_{ADOS}$). Then, using principal component analysis, the first principal component score of CF1 and CF2 (accounting for 86% of the total variance) was taken as a single, parsimonious measure of camouflaging for all subsequent analyses. This measure should be interpreted by relative values (i.e. higher scores indicate more camouflaging) rather than absolute values. This operationalization only allows for estimating camouflaging in autistic individuals in our cohort, as it partly derives from the ADOS score which was not available in TD participants. This approach remains informative, as qualitative studies suggest that camouflaging in autism can be different from similar phenomenon (e.g. impression management) in TD individuals[76,77].

In-vivo chemogenetic manipulation of excitation in mouse prefrontal cortex

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. The Italian Ministry of Health specifically approved the protocol of this study, authorization no. 852/17 to A.G. All surgical procedures were performed under anesthesia.

Six to eight week-old adult male C57Bl6/J mice (Jackson Laboratories; Bar Harbor, ME, USA) were anesthetized with isoflurane (isofluorane 4%,) and head-fixed in a mouse stereotaxic apparatus (isofluorane 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 μl/min, followed by a 5–10 min waiting period, to avoid backflow of viral solution and unspecific labeling. Viral suspensions were injected bilaterally in PFC using the following coordinates, expressed in millimeter from bregma: 1.7 from anterior to posterior, 0.3 lateral, −1.7 deep. The inhibitory DREADD hM4Di 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 0.3 μL injection volume in n=15 hM4Di DREADD and n=19 SHAM mice, respectively. The excitatory DREADD hM3Dq was transduced using an AAV8-CamkII-hM3D (Gq)-mCherry construct. Control animals for this experiment were injected with a control AAV8-hSyn-GFP construct. This set of injection were carried out using a 1μL injection

volume in n=17 hM3Dq DREADD and n=19 SHAM mice, respectively. We waited at least 3 weeks to allow for maximal viral expression.

Mouse rsfMRI data acquisition

The animal preparation protocol for mouse rsfMRI scanning was previously described in detail[78]. Briefly, mice were anesthetized with isoflurane (5% induction), intubated and artificially ventilated (2% maintenance). Then isoflurane was discontinued and substituted with halothane (0.75%), a sedative that preserves cerebral blood flow auto-regulation and neurovascular coupling[79]. Functional data acquisition commenced 30 min after isoflurane cessation. CNO (2 mg/kg for hM4Di and 0.5 mg/kg for hM3Dq) was administered i.v. after 15 minutes from the beginning of the acquisition both in virally transduced animals and in sham mice.

Mouse rsfMRI data analysis

Raw mouse rsfMRI data was preprocessed as described in previous work[80,81]. Briefly, the initial 120 volumes of the time series were removed to allow for T1 and gradient equilibration effects. Data were then despiked, motion corrected and spatially registered to a common reference template. Motion traces of head realignment parameters (3 translations + 3 rotations) and mean ventricular signal (corresponding to the averaged BOLD signal within a reference ventricular mask) were used as nuisance covariates and regressed out from each time course. All rsfMRI time series also underwent band-pass filtering to a frequency window of 0.01–0.1 Hz and spatial smoothing with a full width at half maximum of 0.6 mm.

The experimental design of the study allowed for computation of H during time-windows in the rsfMRI scan before drug injection (e.g., "Baseline"), a transition phase where the drug begins having its effect (e.g., "Transition"), and a treatment phase when the drug is thought to have its optimal effect (e.g., "Treatment"). Analysis of condition, treatment phase, time, and all interactions between such factors was achieved using a sliding window analysis. Each window was 512 volumes in length and the sliding step was 1 volume. H is computed at each window and results in an H time-series. The H time-series is used as the dependent variable in a linear mixed effect model (i.e. using the *lme* function within the *nlme* library in R) with fixed effects of condition, time, treatment phase, and all 2-way and 3-way interactions between such factors as well as a factor accounting for scan day. Random effects in the model included time within mouse as well as treatment phase within mouse, all modeled with random intercepts and slopes. This omnibus model was utilized to examine a 3-way interaction between condition, time, and treatment phase. If this interaction was present, we then split the data by the 3 levels of the treatment phase (e.g., Baseline, Transition, and Treatment), in order to examine the main effect of condition or the condition*time interaction. Plots of the data indicate each mouse (grey lines in Figure 3) as well as group trajectories for each phase, with all trajectories estimated with a generalized additive model smoother applied to individual mice and group trajectories.

In-silico recurrent network modeling of LFP and BOLD data

The recurrent network model we use represents a standard cortical circuit incorporating integrate-and-fire excitatory and inhibitory spiking neurons that interact through recurrent

connections and receive external inputs (both a sensory driven thalamic input and a sensory unrelated intracortical input, see Figure 1A). The network structure and parameters of the recurrent network model are the same ones used in Cavallari et al.,[34] with conductance-based synapses (for full details see[34]). 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: a constant-rate thalamic input and an intracortical input generated by an Ornstein-Uhlenbeck (OU) process with zero mean. A description of the baseline reference parameters used in simulations is given in Table 4. The LFP is computed as the sum of absolute values of AMPA and GABA postsynaptic currents on excitatory cells[29,30]. 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 data from a complex three-dimensional model of the dipoles generated by cortical neurons[29,30,33]. We changed the E:I ratio by independently varying the strengths of the inhibitory (g_I) and excitatory (g_E) synaptic conductances. We called g the relative ratio between inhibitory and excitatory conductances ($g = g_I/g_E$). We present results of simulations for two levels of strength of thalamic input (υ_0 = 1.5 spikes/second and υ_0 = 2 spikes/second), and we verified that our results hold qualitatively for a wider range of input levels (1.5 to 4 spikes/second).

For the simulations used to compute H and 1/f slope of LFPs, we simulated a 10-second stretch of network activity from which we extracted a 10-second LFP time series used to compute H and 1/f slopes for each individual value of g (Figure 1B). To estimate power spectral density (PSD) we computed the Fast Fourier Transform with the Welch's method, dividing the data into ten overlapping segments with 50% overlap. 1/f slopes were computed with least-squares regressions predicting log power with log frequency. A piece-wise regression was applied to fit two line segments to the PSD – one segment to a low frequency region from 1-30 Hz and a second segment to a high-frequency region from 30-100 Hz.

As a basis for our model translating BOLD from LFP data, we note that prior studies with simultaneous electrophysiological and fMRI recordings in non-human primates have established that BOLD signal amplitude is more closely correlated with LFP than with any other type of neuronal events, such as spikes[36,37]. Similarly, simultaneous electroencephalogram (EEG)/fMRI studies in humans have found that the BOLD correlates with the EEG[39,82], which in turns correlates strongly with the LFP[83]. Importantly, the BOLD amplitude at any given time has been found to correlate preferentially with the power of high frequency bands. In particular, the BOLD amplitude correlates strongly with the gamma (40 - 100 Hz) band. However, the power distribution across frequency bands carries complementary information about the BOLD signal, meaning that each band contributes to the prediction of BOLD and predicting the BOLD signal directly from a wide band (i.e., the whole LFP spectrum) leads to poorer predictions of BOLD[36–39,84,85]. To account for these empirical observations, we have developed a model of BOLD signal that integrates contributions from different bands with a preferential contribution from high frequency bands.

To compute the simulated BOLD through the convolutions of the simulated LFPs, we needed to generate longer time series than the initial 10 seconds simulated for LFPs. However, it was unfeasible to simulated very long BOLD time-series due to limitations on computational resources. We thus created, from the LFP data used for evaluations of individual g values,

aggregated LFP time-series corresponding to different intervals of g (rather than individual values of g), as follows. The set of 10-second LFP time series was divided into 3 equi-populated groups of g: $g < 7.5$, $7.5 < g < 11$ and $g > 11$. A concatenated LFP time series was created for each group by randomly concatenating 20 LFP traces, which provided 200-second LFP signals. Low frequencies of the concatenated data were log-log linearly extrapolated based on the low-frequency slopes obtained in LFP log-log linear piecewise fitting. To account for statistical variability, the process of concatenating data was repeated 20 times for each group of g, randomly changing the order in which the individual LFP traces were combined within the group.

Once concatenated LFP time-series data was simulated, we compute the simulated BOLD time series as the LFP data convolved not only with a hemodynamic response function (HRF), as in standard network models[86–88], but also with a high-pass filter (HPF) that gives more predictive power to higher LFP frequencies (Supplementary Figure 3B). We have tested different parameters of the HPF, checking that changing the parameters produce qualitatively similar results and a monotonic correspondence between H of simulated LFP and H of simulated BOLD, and we opted for a HPF with a cutoff frequency (the frequency where the response is lowered by 3 dB) of 12.5 Hz and with a peak response at 20 Hz. The effect of the HPF was to attenuate low frequencies of the BOLD power distribution, partly compensating the low-frequency enhancement of HRFs, and to shift the peak frequency of BOLD power to 0.03 Hz, a value much closer to the peak frequency found in our real BOLD data and in most BOLD studies[89,90] with respect to the one that would have been obtained without convolution (see Supplementary Figure 3).

To simulate BOLD response from the LFP data generated from the recurrent model, within the frequency domain we multiplied the LFP spectrum with spectra of the high-pass filter (HPF) and the hemodynamic response function (HRF), as follows:

$$FFT(BOLD) = FFT(LFP)FFT(HPF)[FFT(HRF) + \eta]$$

Where FFT is the fast Fourier transform operation and η is a constant white noise term. This noise term summarizes neurovascular relationship at frequencies not observable because they are faster than the BOLD acquisition frequencies. We assumed that the amplitude of η was very small and we assigned small values to this noise term. We checked that the exact value of amplitude of η, or the spectral profile of this noise term (for example, simulating 1/f noise instead of white noise), did not alter the monotonic relationship between the simulated H of BOLD and LFP. Finally, the simulated BOLD time-series data was produced by applying the inverse Fourier transform of $FFT(BOLD)$ and then downsampling the resulting signal to a lower sampling rate, similar to that used in BOLD experiments (e.g., 0.5 Hz). The HRF used for simulating BOLD was the HRF from Magri et al.,[36], but similar results were obtained when using a canonical HRF (see Supplementary Figure 4).

Analyses examining enrichment of autism-associated genes in different cell types with genes differentially expressed by androgen hormones

To test hypotheses regarding cell types that may be affected by androgen influence, we examined genes linked to autism via rare de novo protein truncating variants that are enriched for expression in specific cell types[42]. Of the 102 genes reported by Satterstrom et al., we split these

lists by enrichments in early excitatory neurons (C3), MGE derived cortical interneurons (C16), microglia (C19), and astrocytes or oligodendrocyte precursor cells (C4). In addition to high risk mutations linked to autism, we additionally used a list of genes differentially expressed (DE) in different cell types within post-mortem prefrontal and anterior cingulate cortex tissue of autistic patients[43]. These DE gene lists were split into cell types, and we examined DE genes in any excitatory neuronal cell class (L2/3, L4/ L5/6), inhibitory cell classes (IN-PV, IN-SST, IN-VIP, IN-SV2C), microglia, astrocytes (AST-PP, AST-FB), and oligodendrocytes.

To test the question of whether cell type autism-associated gene lists were enriched for genes known to be differentially expressed by DHT, we used a previous DE gene list from an RNA-seq dataset of DHT administration to human neuronal stem cells was utilized for these tests[13]. Custom code was utilized to compute enrichment odds ratios and hypergeometric p-values for each enrichment test with different cell type autism-associated lists. The background total for these tests was the total number of genes considered in the original DHT-administration dataset (13,284).

To test how the DHT-sensitive and autism-associated genes in excitatory neurons are expressed across the human adult brain, we used whole-brain maps of expression for each gene in MNI space from the Allen Institute Human Brain Atlas[45]. Maps for each gene were downloaded from the Neurosynth website (https://neurosynth.org/genes/) and then submitted to a one-sample t-test in SPM12, with a threshold of FDR $q<0.01$.

Data and code availability

Tidy data and analysis code are available at https://github.com/landiit/EI_hurst. Source code of the recurrent network model is available at https://github.com/pablomc88/EEG_proxy_from_network_point_neurons.

Acknowledgments

The Medical Research Council Autism Imaging Multicentre Study Consortium (MRC AIMS Consortium) is a UK collaboration between the Institute of Psychiatry, Psychology & Neuroscience (IoPPN) at King's College London, the Autism Research Centre, University of Cambridge and the Autism Research Group, University of Oxford. The Consortium members are in alphabetical order: Anthony J Bailey (Oxford), Simon Baron-Cohen (Cambridge), Patrick F Bolton (IoPPN), Edward T Bullmore (Cambridge), Sarah Carrington (Oxford), Marco Catani (IoPPN), Bhismadev Chakrabarti (Cambridge), Michael C Craig (IoPPN), Eileen M Daly (IoPPN), Sean CL Deoni (IoPPN), Christine Ecker (IoPPN), Francesca Happé (IoPPN), Julian Henty (Cambridge), Peter Jezzard (Oxford), Patrick Johnston (IoPPN), Derek K Jones (IoPPN), Meng-Chuan Lai (Cambridge), Michael V Lombardo (Cambridge), Anya Madden (IoPPN), Diane Mullins (IoPPN), Clodagh M Murphy (IoPPN), Declan GM Murphy (IoPPN), Greg Pasco (Cambridge), Amber NV Ruigrok (Cambridge), Susan A Sadek (Cambridge), Debbie Spain (IoPPN), Rose Stewart (Oxford), John Suckling (Cambridge), Sally J Wheelwright (Cambridge) and Steven C Williams (IoPPN).

Funding

The authors disclosed receipt of the following financial support for the research, authorship and/or publication of this article: The project was supported by the UK Medical Research Council (grant number: GO 400061) and European Autism Interventions – a Multicentre Study for Developing New Medications (EU-AIMS); EU-AIMS has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115300, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in-kind contribution. During the period of this work, M.-C.L. was supported by the Ontario Brain Institute via the Province of Ontario Neurodevelopmental Disorders Network (IDS-I l-02), the Slifka-Ritvo Award for Innovation in Autism Research from the International Society for Autism Research and the Alan B. Slifka Foundation, the Academic Scholars Award from the Department of Psychiatry, University of Toronto, the O'Brien Scholars Program in the Child and Youth Mental Health Collaborative at the Centre for Addiction and Mental Health (CAMH) and The Hospital for Sick Children, the Slaight Family Child and Youth Mental Health Innovation Fund from CAMH Foundation, and the Canadian Institutes of Health Research; M.V.L. was supported by a European Research Council (ERC) Starting Grant (ERC-2017-STG; AUTISMS; no. 755816); A. G. was supported by a European Research Council (ERC) Starting Grant (ERC-2018-STG; DISCONN; no. 802371), the Brain and Behavior Research Foundation (NARSAD; Independent Investigator Grant; no. 25861), the Simons Foundation (SFARI 400101), the NIH (1R21MH116473-01A1) and by Telethon foundation; S. B.-C. was funded by the Autism Research Trust, the Wellcome Trust, the Templeton World Charitable Foundation, and the NIHR Biomedical Research Centre in Cambridge, during the period of this work. S. B.-C. also received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 777394. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and AUTISM SPEAKS, Autistica, SFARI. S. B.-C. is also supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care East of England at

Cambridgeshire and Peterborough NHS Foundation Trust. The views expressed are those of the author(s) and not necessarily those of the NHS, NIHR or Department of Health and Social Care. SP was funded by a SFARI explorer grant (Grant no. 602849), by the NIH Brain Initiative (Grants No. R01 NS108410 and No. U19 NS107464), and by the CARITRO Foundation.

Contributions

S.P., A.G., M.-C.L., and M.V.L. have equal senior author contributions. All authors, including the MRC AIMS Consortium, contributed to the conception and design of the study, performed the research, contributed to the collection of data, and provided critical inputs in the writing of the manuscript. S.T., P.M.-C., S.P., and M.V.L. analyzed the data. S.T., P.M.-C., M.-C.L., S.P., A.G., and M.V.L. drafted the manuscript.

Declaration of Interests

E.T.B. is employed half-time by the University of Cambridge and half-time at GlaxoSmithKline plc (GSK); he holds stock in GSK. All other authors have no conflict of interests to declare.

References

1. Rubenstein, J. L. R. & Merzenich, M. M. Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes Brain Behav.* **2**, 255–267 (2003).
2. Sohal, V. S. & Rubenstein, J. L. R. Excitation-inhibition balance as a framework for investigating mechanisms in neuropsychiatric disorders. *Mol. Psychiatry* **24**, 1248–1257 (2019).
3. Nelson, S. B. & Valakh, V. Excitatory/Inhibitory Balance and Circuit Homeostasis in Autism Spectrum Disorders. *Neuron* **87**, 684–698 (2015).
4. Happé, F., Ronald, A. & Plomin, R. Time to give up on a single explanation for autism. *Nat. Neurosci.* **9**, 1218–1220 (2006).
5. Lombardo, M. V., Lai, M.-C. & Baron-Cohen, S. Big data approaches to decomposing heterogeneity across the autism spectrum. *Mol. Psychiatry* **24**, 1435–1450 (2019).
6. Lombardo, M. V. *et al.* Different functional neural substrates for good and poor language outcome in autism. *Neuron* **86**, 567–577 (2015).
7. Lombardo, M. V. *et al.* Large-scale associations between the leukocyte transcriptome and BOLD responses to speech differ in autism early language outcome subtypes. *Nat. Neurosci.* **21**, 1680–1688 (2018).
8. Lombardo, M. V. *et al.* Default mode-visual network hypoconnectivity in an autism subtype with pronounced social visual engagement difficulties. *Elife* **8**, (2019).
9. Lai, M.-C., Lombardo, M. V., Auyeung, B., Chakrabarti, B. & Baron-Cohen, S. Sex/gender differences and autism: setting the scene for future research. *J Am Acad Child Adolesc Psychiatry* **54**, 11–24 (2015).
10. Lai, M.-C. *et al.* Imaging sex/gender and autism in the brain: Etiological implications. *J. Neurosci. Res.* **95**, 380–397 (2017).
11. Lee, E., Lee, J. & Kim, E. Excitation/Inhibition Imbalance in Animal Models of Autism Spectrum Disorders. *Biol. Psychiatry* **81**, 838–847 (2017).
12. Bozzi, Y., Provenzano, G. & Casarosa, S. Neurobiological bases of autism-epilepsy comorbidity: a focus on excitation/inhibition imbalance. *Eur. J. Neurosci.* **47**, 534–548 (2018).
13. Lombardo, M. V. *et al.* Sex-specific impact of prenatal androgens on social brain default mode subsystems. *Mol. Psychiatry* (2018) doi:10.1038/s41380-018-0198-y.
14. Yizhar, O. *et al.* Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature* **477**, 171–178 (2011).
15. Selimbeyoglu, A. *et al.* Modulation of prefrontal cortex excitation/inhibition balance rescues social behavior in CNTNAP2-deficient mice. *Sci Transl Med* **9**, (2017).
16. Lai, M.-C. *et al.* Quantifying and exploring camouflaging in men and women with autism. *Autism* **21**, 690–702 (2017).
17. Hull, L. *et al.* Gender differences in self-reported camouflaging in autistic and non-autistic adults. *Autism* 1362361319864804 (2019) doi:10.1177/1362361319864804.
18. Schuck, R. K., Flores, R. E. & Fung, L. K. Brief Report: Sex/Gender Differences in Symptomology and Camouflaging in Adults with Autism Spectrum Disorder. *J Autism Dev Disord* **49**, 2597–2604 (2019).
19. Lai, M.-C. *et al.* Neural self-representation in autistic women and association with 'compensatory camouflaging'. *Autism* **23**, 1210–1223 (2019).

20. Bullmore, E. *et al.* Colored noise and computational inference in neurophysiological (fMRI) time series analysis: Resampling methods in time and wavelet domains. *Human Brain Mapping* **12**, 61–78 (2001).
21. Maxim, V. *et al.* Fractional Gaussian noise, functional MRI and Alzheimer's disease. *Neuroimage* **25**, 141–158 (2005).
22. He, B. J. Scale-free properties of the functional magnetic resonance imaging signal during rest and task. *J. Neurosci.* **31**, 13786–13795 (2011).
23. Bédard, C., Kröger, H. & Destexhe, A. Does the 1/f frequency scaling of brain signals reflect self-organized critical states? *Phys. Rev. Lett.* **97**, 118102 (2006).
24. He, B. J. Scale-free brain activity: past, present, and future. *Trends Cogn. Sci. (Regul. Ed.)* **18**, 480–487 (2014).
25. Gao, R., Peterson, E. J. & Voytek, B. Inferring synaptic excitation/inhibition balance from field potentials. *Neuroimage* **158**, 70–78 (2017).
26. Lai, M.-C. *et al.* A shift to randomness of brain oscillations in people with autism. *Biol. Psychiatry* **68**, 1092–1099 (2010).
27. Watanabe, T., Rees, G. & Masuda, N. Atypical intrinsic neural timescale in autism. *eLife* **8**, e42256 (2019).
28. Stadnitski, T. Measuring fractality. *Front Physiol* **3**, 127 (2012).
29. Mazzoni, A., Panzeri, S., Logothetis, N. K. & Brunel, N. Encoding of naturalistic stimuli by local field potential spectra in networks of excitatory and inhibitory neurons. *PLoS Comput. Biol.* **4**, e1000239 (2008).
30. Mazzoni, A. *et al.* Computing the Local Field Potential (LFP) from Integrate-and-Fire Network Models. *PLoS Comput. Biol.* **11**, e1004584 (2015).
31. Mazzoni, A., Whittingstall, K., Brunel, N., Logothetis, N. K. & Panzeri, S. Understanding the relationships between spike rate and delta/gamma frequency bands of LFPs and EEGs using a local cortical network model. *Neuroimage* **52**, 956–972 (2010).
32. Mazzoni, A., Brunel, N., Cavallari, S., Logothetis, N. K. & Panzeri, S. Cortical dynamics during naturalistic sensory stimulations: experiments and models. *J. Physiol. Paris* **105**, 2–15 (2011).
33. Barbieri, F., Mazzoni, A., Logothetis, N. K., Panzeri, S. & Brunel, N. Stimulus dependence of local field potential spectra: experiment versus theory. *J. Neurosci.* **34**, 14589–14605 (2014).
34. Cavallari, S., Panzeri, S. & Mazzoni, A. Comparison of the dynamics of neural interactions between current-based and conductance-based integrate-and-fire recurrent networks. *Front Neural Circuits* **8**, 12 (2014).
35. Haller, M. *et al.* *Parameterizing neural power spectra*. http://biorxiv.org/lookup/doi/10.1101/299859 (2018) doi:10.1101/299859.
36. Magri, C., Schridde, U., Murayama, Y., Panzeri, S. & Logothetis, N. K. The amplitude and timing of the BOLD signal reflects the relationship between local field potential power at different frequencies. *J. Neurosci.* **32**, 1395–1407 (2012).
37. Logothetis, N. K., Pauls, J., Augath, M., Trinath, T. & Oeltermann, A. Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412**, 150–157 (2001).
38. Schölvinck, M. L., Maier, A., Ye, F. Q., Duyn, J. H. & Leopold, D. A. Neural basis of global resting-state fMRI activity. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10238–10243 (2010).
39. Scheeringa, R. *et al.* Neuronal dynamics underlying high- and low-frequency EEG oscillations contribute independently to the human BOLD signal. *Neuron* **69**, 572–583 (2011).

40. Alexander, G. M. *et al.* Remote control of neuronal activity in transgenic mice expressing evolved G protein-coupled receptors. *Neuron* **63**, 27–39 (2009).
41. Stachniak, T. J., Ghosh, A. & Sternson, S. M. Chemogenetic synaptic silencing of neural circuits localizes a hypothalamus→midbrain pathway for feeding behavior. *Neuron* **82**, 797–808 (2014).
42. Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* (2020) doi:10.1016/j.cell.2019.12.036.
43. Velmeshev, D. *et al.* Single-cell genomics identifies cell type–specific molecular changes in autism. *Science* **364**, 685–689 (2019).
44. Quartier, A. *et al.* Genes and Pathways Regulated by Androgens in Human Neural Cells, Potential Candidates for the Male Excess in Autism Spectrum Disorder. *Biological Psychiatry* **84**, 239–252 (2018).
45. Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).
46. Dhamala, E., Jamison, K. W., Sabuncu, M. R. & Kuceyeski, A. *Sex classification using long-range temporal dependence of resting-state functional MRI time series*. http://biorxiv.org/lookup/doi/10.1101/809954 (2019) doi:10.1101/809954.
47. Courchesne, E. *et al.* The ASD Living Biology: from cell proliferation to clinical phenotype. *Mol. Psychiatry* **24**, 88–107 (2019).
48. Willsey, A. J. *et al.* Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013).
49. Parikshak, N. N. *et al.* Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021 (2013).
50. Di Martino, A. *et al.* Functional brain correlates of social and nonsocial processes in autism spectrum disorders: an activation likelihood estimation meta-analysis. *Biol. Psychiatry* **65**, 63–74 (2009).
51. Di Martino, A. *et al.* The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol Psychiatry* **19**, 659–667 (2014).
52. Uddin, L. Q. & Menon, V. The anterior insula in autism: under-connected and under-examined. *Neurosci Biobehav Rev* **33**, 1198–1203 (2009).
53. Padmanabhan, A., Lynch, C. J., Schaer, M. & Menon, V. The Default Mode Network in Autism. *Biol Psychiatry Cogn Neurosci Neuroimaging* **2**, 476–486 (2017).
54. Lombardo, M. V. *et al.* Atypical neural self-representation in autism. *Brain* **133**, 611–624 (2010).
55. Buckner, R. L. & DiNicola, L. M. The brain's default network: updated anatomy, physiology and evolving insights. *Nat. Rev. Neurosci.* **20**, 593–608 (2019).
56. Yeo, B. T. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).
57. Horder, J. *et al.* GABAA receptor availability is not altered in adults with autism spectrum disorder or in mouse models. *Sci Transl Med* **10**, (2018).
58. Coghlan, S. *et al.* GABA system dysfunction in autism and related disorders: From synapse to symptoms. *Neuroscience & Biobehavioral Reviews* **36**, 2044–2055 (2012).
59. Kaczkurkin, A. N., Raznahan, A. & Satterthwaite, T. D. Sex differences in the developing brain: insights from multimodal neuroimaging. *Neuropsychopharmacology* **44**, 71–85 (2019).

60. Bedford, S. A. *et al.* Large-scale analyses of the relationship between sex, age and intelligence quotient heterogeneity and cortical morphometry in autism spectrum disorder. *Mol. Psychiatry* (2019) doi:10.1038/s41380-019-0420-6.
61. Livingston, L. A., Shah, P. & Happé, F. Compensatory strategies below the behavioural surface in autism: a qualitative study. *Lancet Psychiatry* **6**, 766–777 (2019).
62. Mitchell, J. P., Macrae, C. N. & Banaji, M. R. Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron* **50**, 655–663 (2006).
63. Robinson, E. B., Lichtenstein, P., Anckarsäter, H., Happé, F. & Ronald, A. Examining and interpreting the female protective effect against autistic behavior. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5258–5262 (2013).
64. Werling, D. M. The role of sex-differential biology in risk for autism spectrum disorder. *Biol Sex Differ* **7**, 58 (2016).
65. Lord, C., Rutter, M. & Le Couteur, A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* **24**, 659–685 (1994).
66. Pilling, S., Baron-Cohen, S., Megnin-Viggars, O., Lee, R. & Taylor, C. Recognition, referral, diagnosis, and management of adults with autism: summary of NICE guidance. *BMJ* **344**, (2012).
67. Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J. & Clubley, E. The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J Autism Dev Disord* **31**, 5–17 (2001).
68. Lord, C. *et al.* The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord* **30**, 205–223 (2000).
69. Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y. & Plumb, I. The 'Reading the Mind in the Eyes' Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J Child Psychol Psychiatry* **42**, 241–251 (2001).
70. Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L. & Petersen, S. E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* **59**, 2142–2154 (2012).
71. Cox, R. W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**, 162–173 (1996).
72. Kundu, P., Inati, S. J., Evans, J. W., Luh, W.-M. & Bandettini, P. A. Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *Neuroimage* **60**, 1759–1770 (2012).
73. Patel, A. X. *et al.* A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series. *Neuroimage* **95**, 287–304 (2014).
74. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
75. Krishnan, A., Williams, L. J., McIntosh, A. R. & Abdi, H. Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. *Neuroimage* **56**, 455–475 (2011).
76. Bargiela, S., Steward, R. & Mandy, W. The Experiences of Late-diagnosed Women with Autism Spectrum Conditions: An Investigation of the Female Autism Phenotype. *J Autism Dev Disord* **46**, 3281–3294 (2016).
77. Hull, L. *et al.* 'Putting on My Best Normal': Social Camouflaging in Adults with Autism Spectrum Conditions. *J Autism Dev Disord* **47**, 2519–2534 (2017).

78. Bertero, A. *et al.* Autism-associated 16p11.2 microdeletion impairs prefrontal functional connectivity in mouse and human. *Brain* **141**, 2055–2065 (2018).
79. Gozzi, A. *et al.* A multimodality investigation of cerebral hemodynamics and autoregulation in pharmacological MRI. *Magnetic Resonance Imaging* **25**, 826–833 (2007).
80. Gutierrez-Barragan, D., Basson, M. A., Panzeri, S. & Gozzi, A. Infraslow State Fluctuations Govern Spontaneous fMRI Network Dynamics. *Curr. Biol.* **29**, 2295-2306.e5 (2019).
81. Liska, A., Galbusera, A., Schwarz, A. J. & Gozzi, A. Functional connectivity hubs of the mouse brain. *Neuroimage* **115**, 281–291 (2015).
82. Scheeringa, R. *et al.* Trial-by-trial coupling between EEG and BOLD identifies networks related to alpha and theta EEG power increases during working memory maintenance. *Neuroimage* **44**, 1224–1238 (2009).
83. Whittingstall, K. & Logothetis, N. K. Frequency-band coupling in surface EEG reflects spiking activity in monkey visual cortex. *Neuron* **64**, 281–289 (2009).
84. Goense, J. B. M. & Logothetis, N. K. Neurophysiology of the BOLD fMRI signal in awake monkeys. *Curr. Biol.* **18**, 631–640 (2008).
85. Kilner, J. M., Mattout, J., Henson, R. & Friston, K. J. Hemodynamic correlates of EEG: a heuristic. *Neuroimage* **28**, 280–286 (2005).
86. Wijeakumar, S., Ambrose, J. P., Spencer, J. P. & Curtu, R. Model-based functional neuroimaging using dynamic neural fields: An integrative cognitive neuroscience approach. *J Math Psychol* **76**, 212–235 (2017).
87. Buss, A. T., Wifall, T., Hazeltine, E. & Spencer, J. P. Integrating the behavioral and neural dynamics of response selection in a dual-task paradigm: a dynamic neural field model of Dux et al. (2009). *J Cogn Neurosci* **26**, 334–351 (2014).
88. Deco, G., Rolls, E. T. & Horwitz, B. 'What' and 'where' in visual working memory: a computational neurodynamical perspective for integrating FMRI and single-neuron data. *J Cogn Neurosci* **16**, 683–701 (2004).
89. Alcauter, S. *et al.* Frequency of spontaneous BOLD signal shifts during infancy and correlates with cognitive performance. *Dev Cogn Neurosci* **12**, 40–50 (2015).
90. Allen, E. A. *et al.* A baseline for the multivariate comparison of resting-state networks. *Front Syst Neurosci* **5**, 2 (2011).

Tables

	Time	**Condition (DREADD - SHAM)**	**Time x Condition**
Baseline	*0.82 (0.372)*	*0.81 (0.369)*	*0.36 (0.549)*
Transition	*5.65 (**0.017**)** *	*3.25 (0.081)*	4.94 (**0.026**)*
Treatment	*0.61 (0.433)*	*12.92 (**0.001**)** **	*0.66 (0.418)*

Table 1: Results from DREADD manipulation to enhance excitation. *F-statistics (p-values in parentheses) for main effects of time, condition, and time*condition interaction for each of the 3 phases of the experiment (Baseline, Transition, Treatment). * = p<0.05, ** = p<0.001.*

	Time	**Condition (DREADD - SHAM)**	**Time x Condition**
Baseline	*0.02 (0.876)*	*4.01 (0.054)*	*1.02 (0.137)*
Transition	*0.04* (0.838)	*0.35 (0.561)*	1.36 (0.243)
Treatment	*0.40 (0.533)*	*0.20* (0.673)	*0.10 (0.786)*

Table 2: Results from DREADD manipulation to enhance inhibition. *F-statistics (p-values in parentheses) for main effects of time, condition, and time*condition interaction for each of the 3 phases of the experiment (Baseline, Transition, Treatment). * = p<0.05, ** = p<0.001.*

	TD males (N=29	Autistic males (N=23)	TD females (N=33)	Autistic females (N = 25)	Sex	Diagnosis	Diagnosis * Sex Interactions
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	F-statistic (p-value)	F-statistic (p-value)	F-statistic (p-value)
Age	28.00 (6.42)	27.13 (7.14)	26.99 (5.34)	27.35 (6.79)	0.1489 (0.7004)	0.0320 (0.8585)	0.2508 (0.6176)
VIQ	110.62 (11.53)	114.70 (13.04)	120.30 (10.06)	114.08 (12.79)	**5.3340 (0.02285) ***	0.3837 (0.53696)	**5.1815 (0.02484)** *
PIQ	120.00 (10.21)	114.57 (15.70)	117.39 (9.27)	110.88 (17.43)	1.4972 (0.22382)	**5.5542 (0.02027) ***	0.0453 (0.83188)
FIQ	116.97 (10.69)	116.39 (14.15)	121.45 (8.33)	114.16 (13.82)	0.4863 (0.48713)	3.3889 (0.06844)	2.2417 (0.13731)
Camouflaging Score (z-values)	-	-0.16 (0.38)	-	0.15 (0.34)		-	-
AQ	15.28 (6.99)	32.70(8.47)	11.97 (4.93)	38.44 (6.34)	0.2606 (0.61074)	300.5921 (2.2e-16)	12.4871 (0.0006)
ADI-R Reciprocal-Social-Interaction	-	17.26 (4.77)	-	16.56 (4.52)	-	-	-
ADI-R Communication	-	14.83 (3.50)	-	13.40 (3.96)	-	-	-

ADI-R Repetition	-	5.17 (2.35)	-	4.24 (1.61)	-	-	-
ADOS Communication	-	3.30 (1.74)	-	1.24 (1.30)	-	-	-
ADOS Social	-	5.48 (3.45)	-	3.48 (3.06)	-	-	-
ADOS Repetition	-	1.09 (1.12)	-	4.30 (1.61)	-	-	-
ADOS Communication + Social Total	-	8.83 (4.87)	-	4.72 (4.09)	-	-	-
RMET	27.14 (3.59)	20.83 (6.87)	28.91(2.35)	22.84 (6.40)	3.9353 (0.04989)	42.3091 (2.704e-9)	0.0166 (0.89769)
Mean FD	0.17 (0.05)	0.20 (0.07)	0.18(0.06)	0.04 (0.17)	0.5184 (0.4731)	1.7716 (0.1860)	1.1041 (0.2958)

Note: TD: Typically Developing; VIQ: verbal IQ; PIQ: performance IQ; FIQ: full-scale IQ ADI-R: Autism Diagnostic Interview–Revised; ADOS: Autism Diagnostic Observation Schedule; AQ: Autism Spectrum Quotient; RMET: Reading the Mind in the Eyes Test; FD: frame-wise displacement; *: $p < 0.05$.

Table 3. *Descriptive statistics and group comparisons for the various demographic variables.*

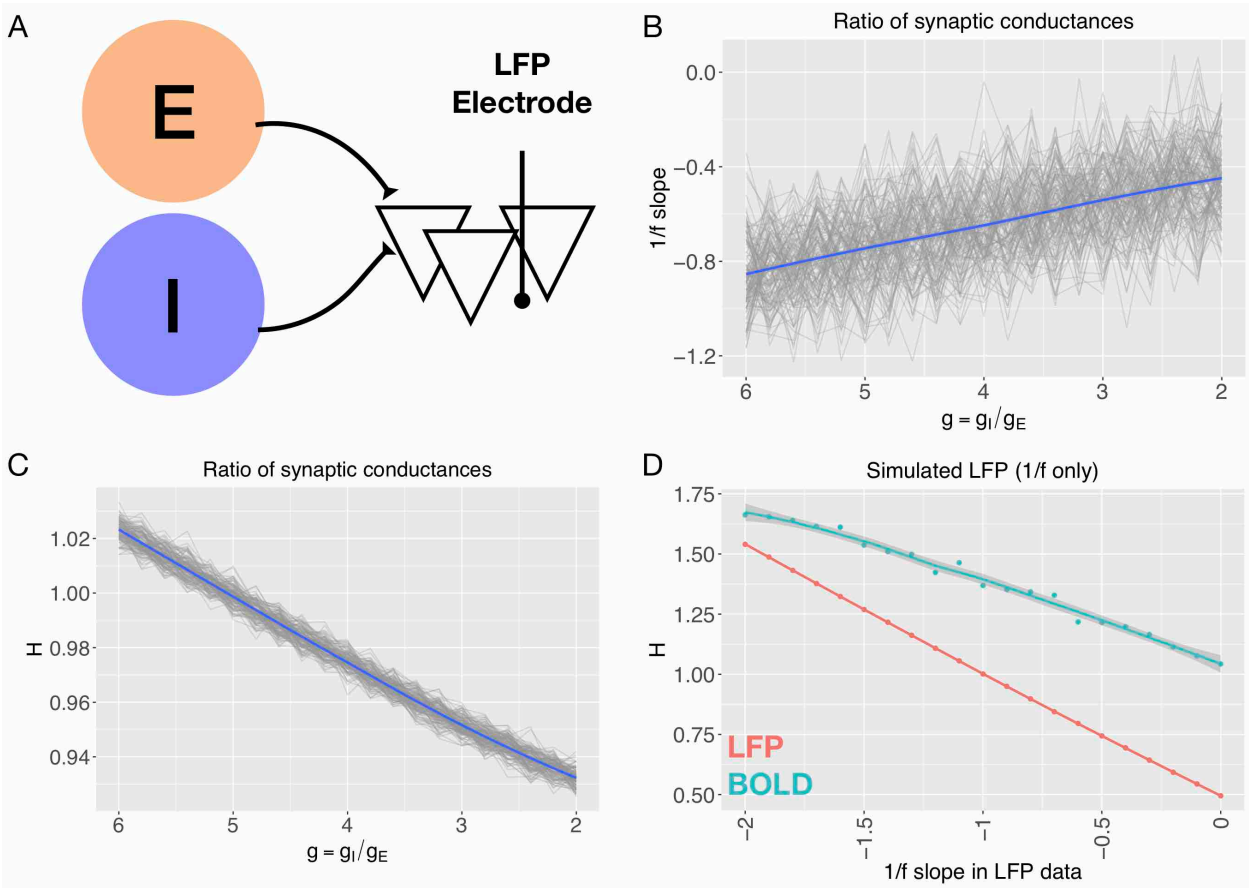
A: Neuron model			
Parameter	**Description**	**Excitatory cells**	**Inhibitory cells**
V_{leak} (mV)	Leak membrane potential	-70	-70
$V_{threshold}$ (mV)	Spike threshold	-52	-52
V_{reset} (mV)	Reset potential	-59	-59
$\tau_{refractory}$ (ms)	Absolute refractory period	2	1
g_{leak} (nS)	Leak membrane conductance	25	20
C_m (pF)	Membrane capacitance	500	200
τ_m (ms)	Membrane time constant	20	10
B: Connection parameters			
Parameter	**Description**	**Excitatory cells**	**Inhibitory cells**
E_{AMPA} (mV)	AMPA reversal potential	0	0
E_{GABA} (mV)	GABA reversal potential	-80	-80
$\tau_{r(AMPA)}$ (ms)	Conductance rise time (AMPA)	0.4	0.2
$\tau_{d(AMPA)}$ (ms)	Conductance decay time (AMPA)	2	1
$\tau_{r(GABA)}$ (ms)	Conductance rise time (GABA)	0.25	0.25
$\tau_{d(GABA)}$ (ms)	Conductance decay time (GABA)	5	5
τ_l (ms)	Synapse latency	0	0
$g_{AMPA(rec.)}$ (nS)	AMPA conductance (recurrent)	0.178	0.233
$g_{AMPA(tha.)}$ (nS)	AMPA conductance (thalamic)	0.234	0.317
$g_{AMPA(cort.)}$ (nS)	AMPA conductance (intracortical)	0.187	0.254
g_{GABA} (nS)	GABA conductance	2.01	2.7

Table 4: Baseline reference parameters of the recurrent network model. *Parameters used in Cavallari et al.,*[34] *with conductance-based synapses.*

Supplementary Information

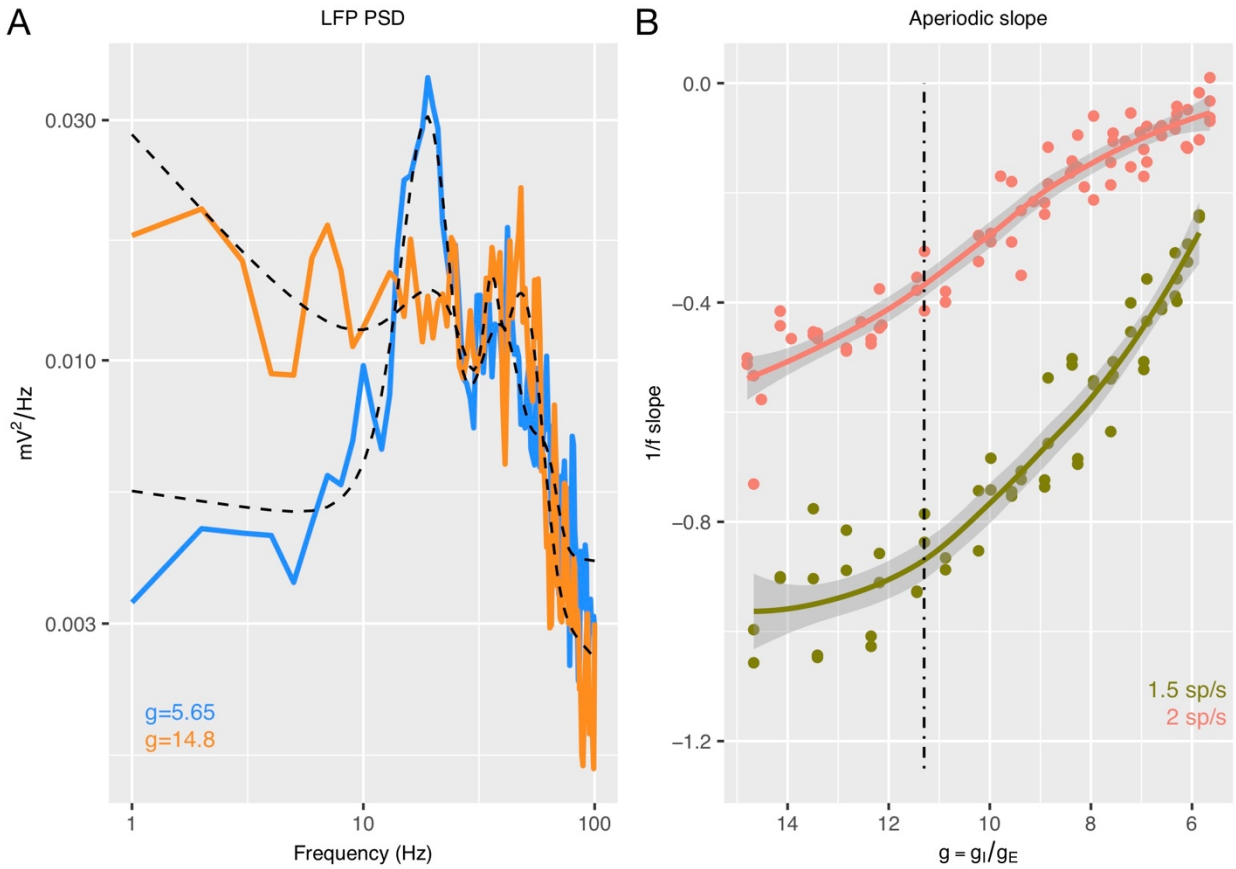
Supplementary Figure 1

In-silico predictions from a non-recurrent model of how 1/f slope and H change with E:I ratio (g). Panel A shows a schematic of the Gao et al. model[25] that computationally generates LFPs based on manipulated E:I ratios. For parameters utilized in this model, see Supplementary Table 1. Panel B reproduces the relationship reported in Gao et al., between 1/f slope and E:I ratio. All 128 simulated channels are shown in gray, and their average is shown in blue. Panel C shows the relationship between H and E:I ratio from the Gao et al., model. Panel D shows H computed on simulated LFP where the 1/f slope is prespecified. H computed on LFP data is shown in pink. The simulated BOLD signal (simulated in the same way as data from the recurrent model) is plotted in cyan.



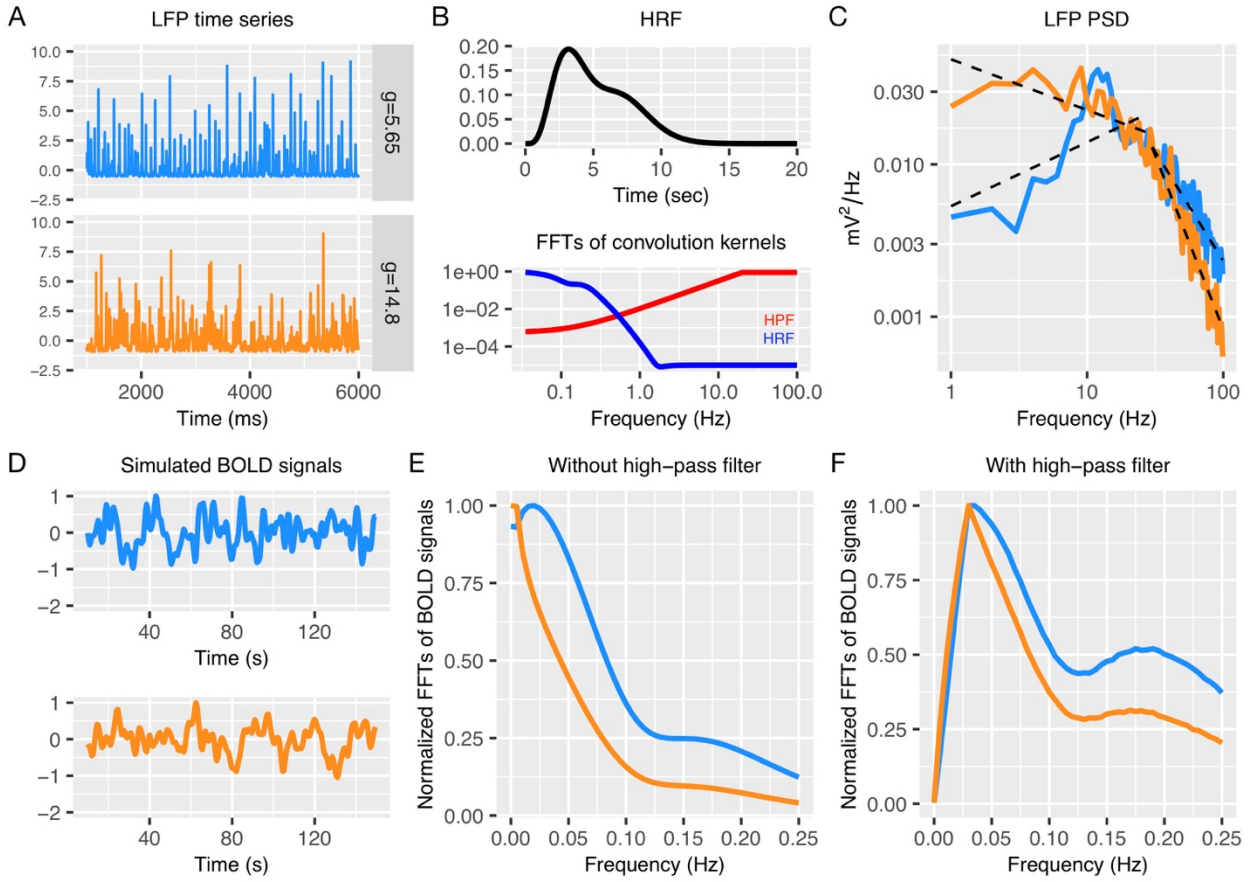
Supplementary Figure 2

Panel A provides two examples of the LFP PSD generated from the recurrent model for different values of the E:I ratio or g (blue, g = 5.65; orange, g = 14.8). Plotted over the PSDs are the FOOOF algorithm[35] model fit (dotted lines), which captures both the aperiodic signal along with the periodic oscillatory signals. Panel B shows how 1/f slope computed from FOOOF changes as a function of g for two different types of thalamic input.



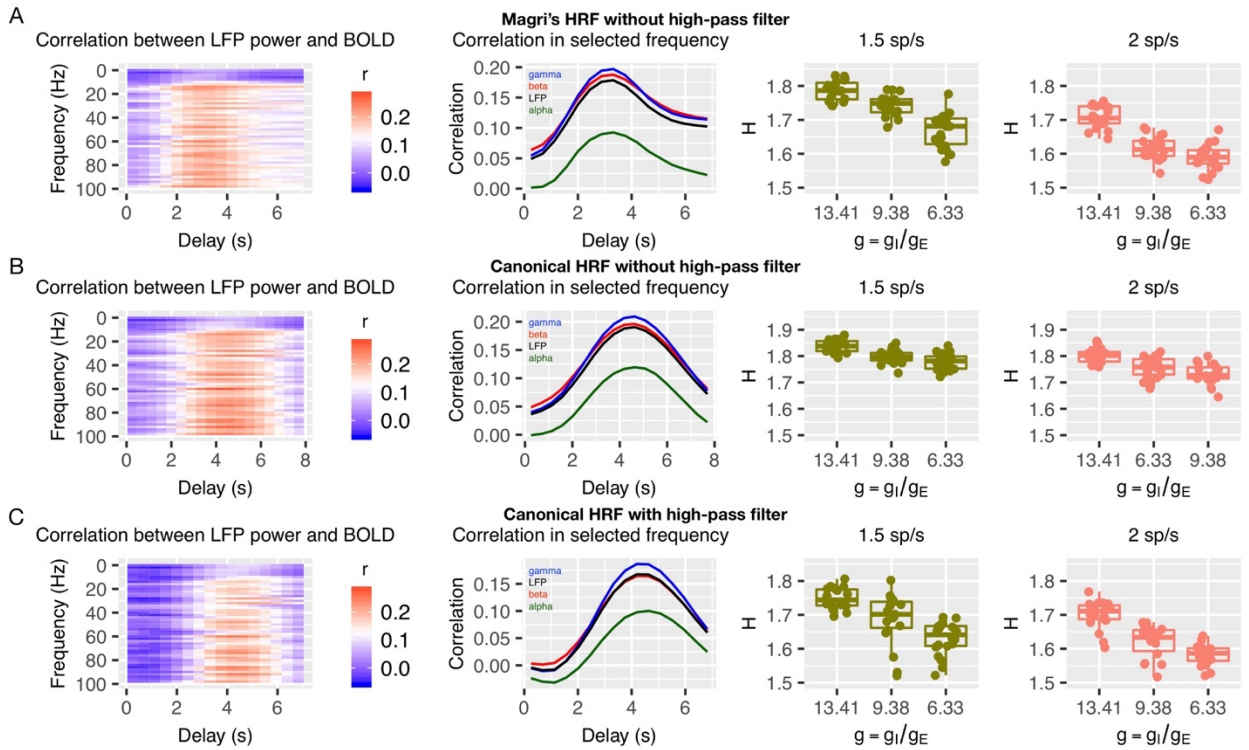
Supplementary Figure 3: Simulating BOLD signal from LFP data generated by the recurrent model.

Panel A shows two example LFP time-series for different values of E:I ratio or g (blue, g = 5.65; orange, g = 14.8). Panel B shows the hemodynamic response function (HRF) from Magri et al.,[36] *(top) and the PSDs from Fast Fourier Transforms (FFT) (bottom) on the HRF (blue) and a high-pass filter (HPF) (red). Panel C shows LFP PSDs. BOLD signal was simulated from LFP data as the result of an inverse FFT of the product of multiplied HRF, HPF, and LFP spectra: FFT(BOLD) = FFT(LFP)FFT(HPF)[FFT(HRF) + η], where η is a constant white noise term. Panel D shows the simulated BOLD signals from this process. Panels E and F show PSD when high-pass filter was or was not used. In particular, BOLD spectra in Panel F show the distinctive properties such as peak frequency (≈ 0.03 Hz) and low frequency to high frequency power ratio (from 1.5 to 4) observed in previous resting-state fMRI studies*[89,90].



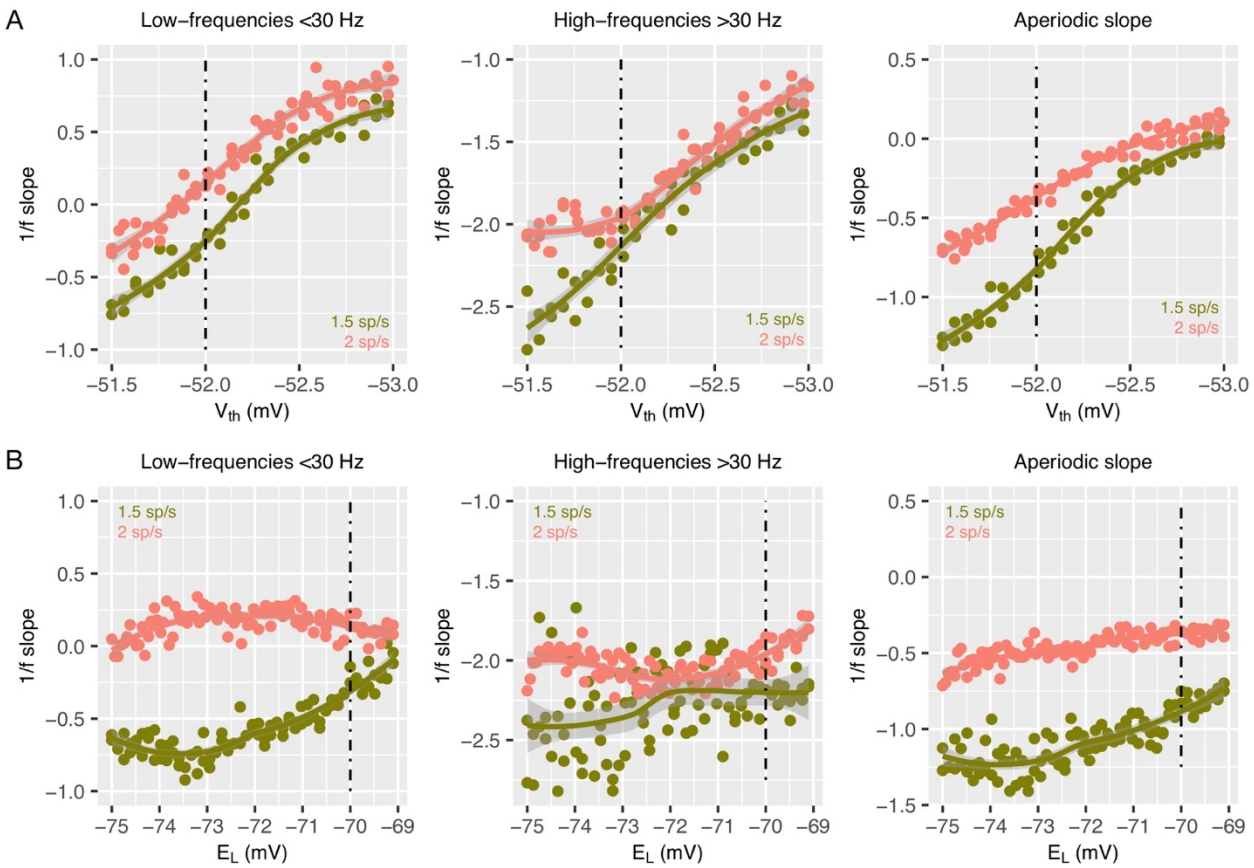
Supplementary Figure 4

Panel A shows a situation where the simulation of BOLD from LFP utilizes the HRF from Magri et al.,[36] *but does not incorporate use of the high-pass filter. Panels B shows BOLD simulated from LFP with the canonical HRF without (B) and with (C) the high-pass filter. In the far left column are time-lagged correlations between LFP power over a range of frequencies as a function of the time-delay. The middle column shows LFP power correlations with BOLD in selected frequency bands. The far right column shows plots for how g is related to H in BOLD for each situation.*



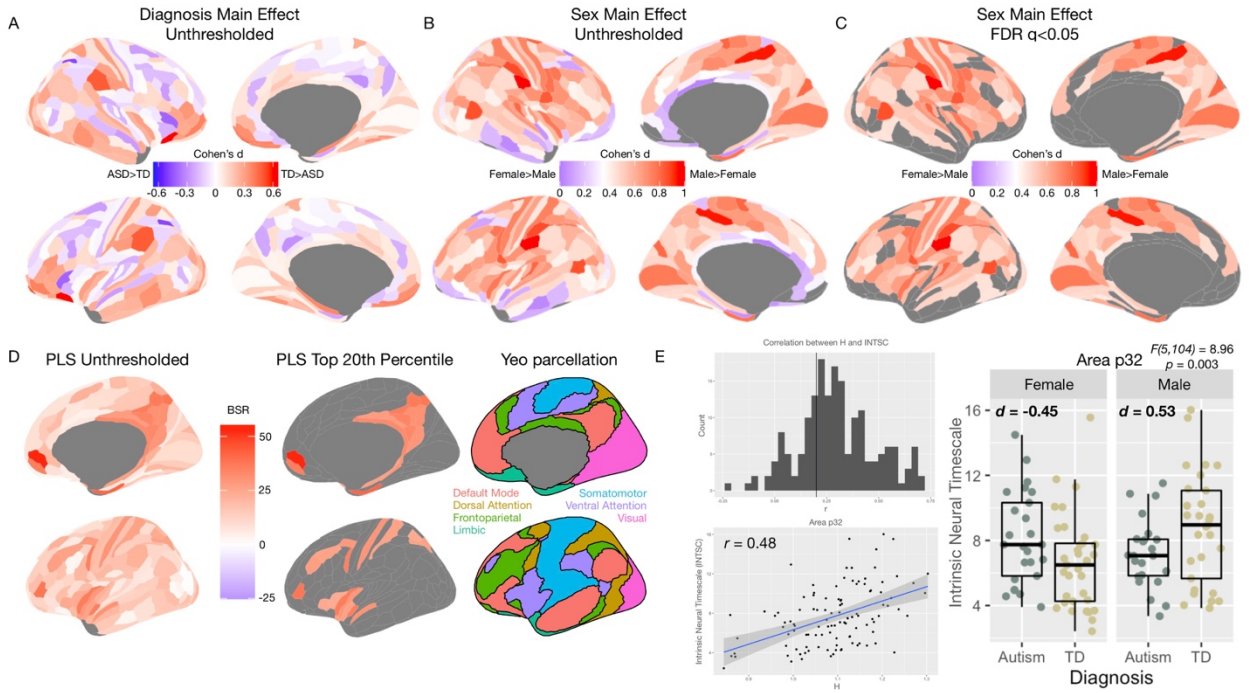
Supplementary Figure 5

Panel A shows 1/f slopes computed after the voltage threshold for spike initiation (V_{th}) is manipulated. Panel B shows 1/f slopes computed after the resting potential for excitatory and inhibitory neurons (E_L) is manipulated. The left column shows results when computed in the low-frequency range (<30 Hz). The middle column shows results when computed in the high-frequency range (>30 Hz). The right column shows 1/f slopes as computed by the FOOOF algorithm[35]. In each plot, salmon color indicates simulations with a thalamic input of 2 spikes/second, while the olive color indicates thalamic input of 1.5 spikes/second.



Supplementary Figure 6

*Panel A shows the unthresholded standardized effect size map for the main effect of Diagnosis in human rsfMRI data from the MRC AIMS dataset. No regions survive at FDR q<0.05 for this comparison. Panel B shows a similar unthresholded standardized effect size map for the main effect of Sex. Panel C shows that several regions (61% of regions) pass at FDR q<0.05 and all are in the directionality of Male>Female. Panel D shows the PLS results next to the 7 cluster rsfMRI parcellation from Yeo et al.,[56]. This panel shows prominent overlap between most important PLS regions and the default mode network (DMN). Panel E shows a histogram of the correlation between H and the intrinsic neural timescale (INTSC) metric reported by Watanabe et al.,[27] across all brain regions (top left). In the bottom left of panel E is a scatterplot showing the correlation between H and INTSC for vMPFC (Area p32). The far right of panel E is a plot to descriptively show the distributions and effect sizes for the sex*diagnosis interaction effect in area p32 for INTSC.*



Parameter	Value
Population Firing Rate (E, I)	2 Hz, 5 Hz
Population Size (E, I)	8000, 2000
Resting Membrane Potential	-65 mV
Reversal Potential (AMPA, $GABA_A$)	0 mV, -80 mV
Conductance Rise Time (AMPA, $GABA_A$)	0.1 ms, 0.5 ms
Conductance Decay Time (AMPA, $GABA_A$)	2 ms, 10 ms
E:I Ratio	1:2 to 1:6

Supplementary Table 1: Parameters for E:I model: *Parameters utilized in the E:I model from Gao et al.,*[25]*, for simulating LFP data.*